Technology Policy and
Assessment Center

Emerging Technologies: Quantitative Identification and Measurement

A Report Prepared for

The Korea Institute of Science and Technology Information

by Susan Cozzens, Sonia Gatchair, Kyung-Sup Kim,
Gonzalo Ordóñez, and Alan Porter

October 29, 2005

Technology Policy and Assessment Center
School of Public Policy, Georgia Institute of Technology
Atlanta GA 30332-034
Phone: 404-894-6703   Fax 404-385-0504
www.tpac.gatech.edu

TABLE OF CONTENTS

TABLE OF CONTENTS *2*

Executive Summary *4*

Chapter One: Emerging Technologies and Innovation Concepts *5*

Emerging Technologies in Innovation and Technological Change **6**

Technological Change, Technological Progress, and Technological Convergence. **11**

Radical Innovation, Disruptive Technology, and Technology Cycle. **11**

Path-Dependence, Lock-ins, and Scientific Revolutions **14**

Strategic Emerging Technologies **14**

Chapter Two: Emerging Technologies and Related Terms *16*

Definitions of Emerging Technologies **16**

Other Candidate Terms for Emerging Technologies **19**

Disruptive Technologies 19

High Technology 20

Relationship between Terms Potentially Related to Emerging Technologies **22**

Chapter Three: Methods for Identifying Emerging Technologies *24*

Qualitative methods **24**

Delphi and Critical Technologies 24

Foresight 25

Data sources **27**

Quantitative techniques **29**

Indexing Categories and Vocabularies 29

Data Mining 30

Combinations 31

Three examples **32**

Predicting emerging technologies with the aid of text-based data mining (Smallheiser, 2001) 32

The emergence of new technologies in the ICT field (Carrocher, 2003) 32

Visualizing science by citation mapping (Small, 1999) 34

Chapter Four: Analyses of Emerging Technologies Using VantagePoint *36*

"Tech Mining" Methodology **36**

Intelligence phase **38**

Data Sources 38

Query Formulation 39

Data Collection and Cleaning 40

Data Summary and Description 41

Design and Analysis Phase **42**

Basic Analyses 42

Advanced Analyses 44

Choice Phase **44**

Chapter Five : Examples of Emerging Technology Applications using VantagePoint 45

Analysis of Emerging Technologies Class Projects **45**
Bio Opportunities for Georgia Tech **46**
Biochips for the Army **46**
Synthetic Lumber **46**
Emerging Technologies for Georgia Tech High Tech Indicators-HTI **46**
Ceramics for Automotive Engine Application **50**
Recent GT Nano Analyses **57**
Discussion and Conclusions **62**
References ***63***
Appendix ***67***
Summary of Data Sources Typically Used in *VantagePoint* **67**
Tech Mining Questions and Sample Indicators **67**

EXECUTIVE SUMMARY

Emerging Technologies: Quantitative Identification and Measurement

Nations increasingly depend on technological competence for growth and prosperity. Emerging technologies present both threats and opportunities to national technology strategies: threats to existing competitive advantages and opportunities to take the lead in new areas before competition becomes entrenched. National governments may therefore find it useful to monitor the global technological frontier for changes that carry significance to their country's technological position.

This report reviews the concepts and methods available to carry out such a monitoring activity systematically, through quantitative methods applied to literature-based data (a set of techniques called bibliometrics).

The concept of emerging technologies is often used but seldom defined formally. We begin by placing the concept into the context of contemporary innovation theory. Based on an analysis of almost 2000 articles that refer to the topic, we identify the core of the concept of emerging technologies as rapid growth, newness, untapped market potential, and a high-technology base.

We turn next to the data and methods available for identifying emerging technologies systematically. There are three major literature-based (or bibliometric) data sources: proposals, publications, and patents. Proposals might give the earliest signals of emerging patterns, but no international data bases are available and lack of consistency in data elements may present a problem for analysis. Publication data bases are international and have the richest set of data elements available for analysis. But it is difficult to distinguish emerging technologies from emerging scientific areas in publication data, and the data itself carries little information on market relevance. Patent data is technologically relevant and yields some information on market potential as viewed by firms, but lags the other two data sources in timing.

To search in these data sources for emerging technologies, the analyst can use either existing indexing categories or data mining techniques that detect structures inherent in the data. The latter approach is more useful for identifying emerging areas, through either cluster or factor analytic techniques. Only co-citation analysis has been applied to large scale data sets to find emerging structures.

Vantage Point is a data mining software package, developed in part at the Technology Policy and Assessment Center at the Georgia Institute of Technology. It has been used to characterize the growth and internal evolution of emerging technologies. Within areas of science or technology, its analytic capabilities can be used with any of the three kinds of bibliometric data to identify and track emerging themes and to summarize indicators of industrial interest.

In summary, bibliometric techniques hold significant potential for use in monitoring systems, but also many limitations. At the current time, national monitoring for emerging technologies therefore must still depend on a combination of quantitative analysis with the knowledge of market-oriented technical experts.

CHAPTER ONE: EMERGING TECHNOLOGIES AND INNOVATION CONCEPTS

Nations increasingly depend on technological competence for growth and prosperity. Today, every sector of a national technology is subject to technological change. Agriculture, even subsistence agriculture, rests on a base of knowledge and technique that is constantly developed and updated through world agricultural research; and biotechnology is transforming competition in specific internationally traded crops. Mineral exploration depends increasingly on sophisticated methods of extraction, and service industries are under constant pressure to incorporate more efficient processes using information technologies. Finally, manufacturing is in a state of continuous creative destruction through the invention and exploitation of new techniques, with nanotechnology looming as the next revolution.

Emerging technologies present both threats and opportunities to national technology strategies. They are threats to existing competitive advantages based on current technological competencies. Even incremental innovations can move the lead from one country or region to another, and radical innovations may eliminate whole markets. But at the same time, competition is more open in new technological areas, and they thus present opportunities to take the lead before competition becomes entrenched.

National governments may therefore find it useful to monitor the global technological frontier for changes that may be important to their country's technological position. Equipped with a list of technological areas where local industry is already competing effectively, national governments may undertake to monitor incremental change among competitors. Even more important, however, is staying attuned to emerging areas that could affect markets for local products but which do not attract the attention of local industry. In theory, then, monitoring emerging technologies internationally and across all of science and engineering is a task national governments might want to sponsor.

The most common approach to identifying emerging technologies is qualitative and expert-based. Even with enhancements designed to combine the strengths of various experts, however, such techniques are prone to gaps – even more so in smaller countries with more narrowly-based expertise. Quantitative approaches may be able to overcome these limitations. This report reviews the concepts and methods available to carry out broad-based monitoring of emerging technological areas systematically, through quantitative methods applied to literature-based data (a set of techniques called bibliometrics).

The concept of emerging technologies is often used but seldom defined formally. The chapter begins by placing the concept into the context of contemporary innovation theory, reviewing a family of basic concepts and relationships related to innovation and technological change. The chapter continues with a characterization of the literature that uses the term. Even without a definition, the literature points in a common direction.

Following that direction, we identify the core of the concept for the purposes of this report as rapid growth, newness, untapped market potential, and a high-technology base.

Emerging Technologies in Innovation and Technological Change

According to a set of definitions adapted in particular from Pavitt (Pavitt, 1991), Nelson & Rosenberg (Nelson & Rosenberg, 1993), and Brooks (Brooks, 1994)[1], "**technology**" can be defined as:

- The tools and machines that help to solve problems (such as a pen or a dam);
- The techniques (know-how) that includes methods, materials, tools, and processes for solving a problem (such as building technology or medical technology);
- A culture-forming activity (such as manufacturing technology, infrastructure technology, or space-travel technology);
- The application of resources to solve a problem (such as knowledge, skills, processes, techniques, tools and raw materials);
- An encompassing term to describe the level of achievement in science, mathematics and engineering of a group or culture.
- The current state of our knowledge of how to combine resources to produce desired products (and our knowledge of what can be produced).
- An activity that either precedes or antecedes both science and engineering.

On the other hand, "**science**" is defined as:

- A reasoned investigation or study of nature and society, aimed at finding out the truth;
- The organized body of knowledge or systems of knowledge gained by such research;
- The intellectual process leading to the development of theories or the application of such theories to increase our understanding of the world and its dynamics.
- It covers any systematic field of study or the knowledge gained from it.
- It comprises the development of concepts, typologies, classifications, models, hypotheses, theories, laws, frameworks, methods, and data.

The difference between Science and Technology is summarized in Table 1 as adapted from Allen (Allen, 1977) and Pavitt (Ibid.). The table shows the difference between these two concepts based on the main purpose of the knowledge being produced, the inputs used, the core activities performed through each one, the type of knowledge result, and the type of problem/need targeted.

[1] http://en.wikipedia.org

Table 1: Differences between Science and Technology

	Science	Technology
Main Purpose	Production of generalizable and reproducible knowledge.	Production of commercializable knowledge embedded in products or services.
Inputs	One or few disciplines. Verbally encoded information.	Multiple disciplines. Verbally encoded information.
Core Activities	Research.	Development, trail, design, experimentation, testing, quality control.
Knowledge Produced	Generalizable, reproducible, codified, publishable.	Specific, partially codified, partially tacit. Disseminated mostly by physical interaction.
Problem/Need Targeted	Mostly diffuse	Mostly specific.

Linked to these concepts, the notion of "**innovation**" is generally used to refer to the idea that both scientific and technological knowledge can have specific strategic value when it is intended to increase market power. More formally, based on the works of Schumpeter in the 30s and 40s, current literature broadly defines innovation as the implementation of a new or significantly improved idea, good, service, process or practice that is intended to be useful and add value to economic activity (OECD, 1997b). Contrary to the notion of technology, the concept of innovation implies a focus on strategic competition and technology use or adoption (Rogers, 2003). Today it is widely accepted that an innovation is not simply a new product, but a whole new market category. Current literature identifies four types of innovations: product innovation, process innovation, organizational innovation, and marketing innovation (Teece, 2002).

Invention, that is, the creation of a new idea or concept, is distinguished from innovation, which implies taking that idea, reducing it to practice, and making it a commercial success. While invention parallels the concept of science, innovation parallels the concept of technology.

How are these concepts related one to each other? According to Brooks (Brooks, 1994), science, technology and innovation although representing a successively larger category of distinct activities, are highly interdependent. As the author posits, science contributes to technology in at least six ways: (1) new knowledge which serves as a direct source of ideas for new technological possibilities; (2) source of tools and techniques for more efficient engineering design and a knowledge base for evaluation of feasibility of designs; (3) research instrumentation, laboratory techniques and analytical methods used in research that eventually find their way into design or industrial practices, often through intermediate disciplines; (4) practice of research as a source for development and assimilation of new human skills and capabilities eventually useful for technology; (5)

creation of a knowledge base that becomes increasingly important in the assessment of technology in terms of its wider social and environmental impacts; (6) knowledge base that enables more efficient strategies of applied research, development, and refinement of new technologies.

The converse impact of technology on science is, as the author claims, of at least equal importance: it impacts (1) through providing a fertile source of novel scientific questions and thereby also helping to justify the allocation of resources needed to address these questions in an efficient and timely manner, extending the agenda of science; and (2) as a source of otherwise unavailable instrumentation and techniques needed to address novel and more difficult scientific questions more efficiently. Furthermore, recent studies based on patent analyses have shown strong relationship between scientific research and innovation (Lim, 2004; Narin, Hamilton, & Olivastro, 1995; Tijssen, 2002). This, however, is still a matter of debate (Meyer, 2000).

In general, many authors find it misleading to take these concepts as meaning different things and following different dynamics. Narin et al. (Narin et al., 1995), for example, argue that based on the new paradigms of knowledge production from the traditional “Mode 1” to “**Mode 2**” which gives more emphasis on application, transdisciplinarity, and diversity of actors involved (Gibbons et al., 1994), together with the fact that engineers are increasingly interested in new and heterodox ways of approaching practical problems, the differences among science, technology, and technological innovation is becoming rather blurred. As the authors claim, modern technology is increasingly science-based. An example of this type of “technoscience” is nanotechnology and biotechnology, where the components of both science and technology are indistinguishable one from the other. A similar discussion relates to the judged inaccurate difference often made between basic and applied research (Stokes, 1997).

In practice, one of the most common ways scholars, managers, and policymakers perceive the relationships among Science, Technology and Innovation is by assuming a **linear sequence**, that is, the presupposition that either science (first basic research, then applied research) leads to new technologies, which in turn leads to innovation, and social and economic outcomes (called the ‘push’ perspective shown in Figure 1) (Bush, 1945), or that the driving force is to be found on the technologies developed, or on the market as a source of new scientific ideas (called the ‘pull’ perspective, which could be represented by inverting the direction of the arrows in Figure 1) (Kline & Rosenberg, 1986; Tijssen, 2002).

Figure 1: The Linear Model of Innovation

Research

⇓

Development

⇓

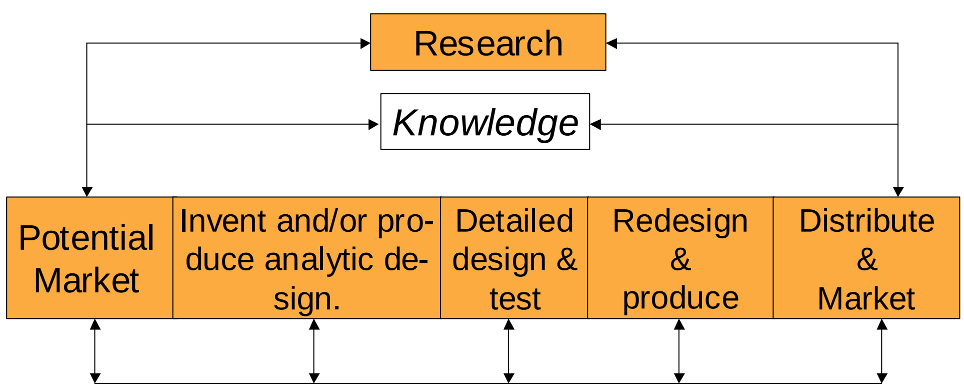
Production

⇓

Marketing

In reaction to this 'oversimplifying' picture of the process of innovation, scholars have proposed a more 'realistic' model of the interaction between research, knowledge, processes, routines, goals and drivers among the relevant factors that influence innovation. Kline and Rosenberg (1986) for example, propose the so called "**chain-link model**" of innovation (Kline & Rosenberg, 1986). According to this approach, the ideas for innovation can steam from many sources, including new manufacturing capabilities, and recognition of market needs. Innovation requires considerable communication among different actors –firms, laboratories, academic institutions and consumers –as well as feedback between science, engineering, product development manufacturing, and marketing. Figure 2 summarizes this model.

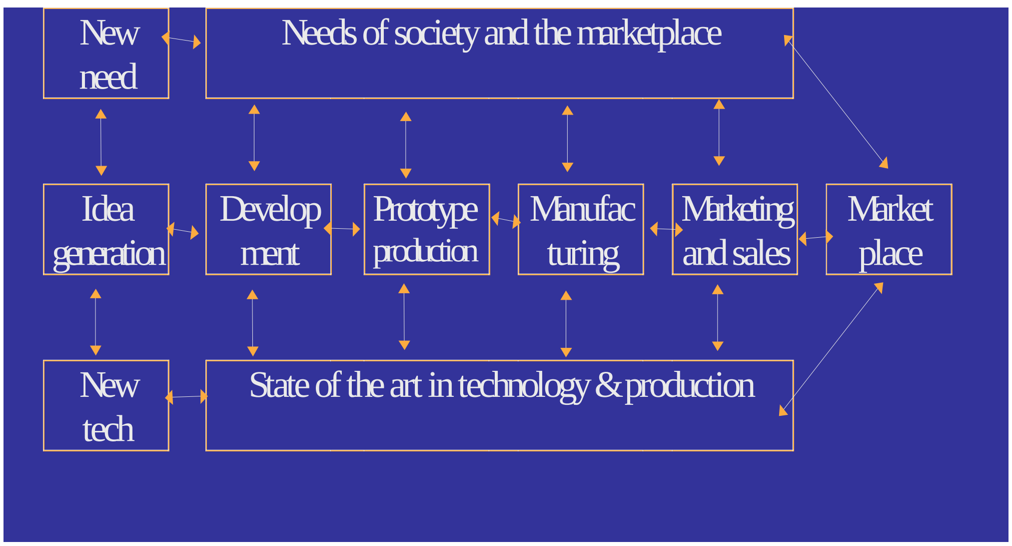
Figure 2: The Chain-Linked Model of Innovation.



More recently, based on the concept of "**innovation systems**", some authors emphasize the socio-technical aspects involving the innovation process. To this scholarly, science, technology, and innovation are shaped, at least partially, by society in general and by the environment and its institutional framework in particular (Carlsson, Jacobsson, Holmen, & Rickne, 2002; Edquist & Hommen, 1999; Etzkowitz & Leydesdorff, 2000; Landry, Amara, & Lamari, 2002; Lundvall, Johnson, Andersen, & Dalum, 2002; Nelson & Rosenberg, 1993; OECD, 1997a). The innovation process, and the way science and technology contributes to it, is portrayed by this approach as the product of the interaction of different actors including those not directly involved in the innovation process. Hence, during the innovation process many players are seen to affect the odds of making a single innovation successful. These include the inventor, someone to fund the development (a venture capitalist or company), suppliers, customers, regulators, lawyers, patent agents, skilled trained individuals, accountants, stock markets, the entrepreneur or innovator, etc.

Figure 3 depicts the main elements of this approach by representing a more complex model of innovation than those described by the linear model and the chain-linked model. In addition to pointing to a variety of sources, the **systemic model** outlines the strategic nature embedded in the technological innovation process: the fact that in practice the purpose of innovation is to meet social, technological, and economic demands as they are presented in society.

Figure 3: A systemic Model of Innovation.



Finally, but not less important, it is also worth reviewing here the concept of "**technology diffusion**". The reason why this notion is important in the framework of this project is that the concept of innovation, by definition, implies use. The need for identifying Emerging Technologies is justified in part by the need for identifying successful innovations, which in turn are measured by the level of adoption. According to Everett Rogers, technology diffusion is the process by which a new idea or new product is accepted by the market. As the author claims, people have different levels of readiness for adopting new innovations, and the characteristics of adoption affect overall adoption. Hence, according to the author, the rate of diffusion is influenced by a) the product's perceived advantage on benefit; b) riskiness of purchase; c) ease of product use (complexity of the product); d) immediacy of benefits; e) observability; f) trialability; g) price; h) extent of behavioral changes required; i) return on investment in the case of industrial products.

Technological Change, Technological Progress, and Technological Convergence.

Associated with the notions of science, technology, and innovation, innovation scholars put forward the concept of "**technological change**" to designate the change in a production function that alters the relationship between inputs and outputs; it is normally understood to be an improvement in technology. In this framework, "**technological progress**" is the consequence of an increase of output resulting from technological change. According to Schumpeter (Schumpeter, 1942) technological progress results as a sudden, punctuated event (creative destruction) resulting only intermittently, caused by a radical innovation that alters the status quo and takes the market to a new equilibrium. In contrast, evolutionary theorists, suggest that technological progress happens as a smooth process of incremental change resulting from learning trajectories followed by innovators and users in a continuous way over time (Nelson & Winter, 1977).

Asymmetric technological progress leads to what is commonly called "**technology gaps**", that is, situations in which some firms or countries benefit from an exclusive set of technologies that allows them to be more efficient and competitive than their competitors. Other ways of identifying technology gaps is when in a certain production chain a technology is missing to make the process more efficient.

As a result of technological change, but not necessarily of technological progress, "**technological convergence**" can be achieved. In current literature, one can identify three types of technological convergence. In one way, having countries as the unit of analysis, technological convergence results when lagged countries increase their technological capabilities and become potential competitors of leading countries through the use of high technologies in the production process. A second type of technological convergence occurs when a variety of technologies are capable of performing competing, similar tasks and although the forms of technology are all very different, they all essentially provide the same basic service. For example, thanks to modern technologies, today one can communicate with anyone through a variety of alternative ways. A third level of technological convergence is when previously separate technologies aiming at different services are integrated to provide multiple services while sharing resources and interacting in such a way that new efficiencies can be created. An example of this is the modern cell phones, which offer the possibility to listen to music, watch movies, serve as an electronic agenda, take pictures, send and receive e-mails, etc. when the owner chooses so.

Radical Innovation, Disruptive Technology, and Technology Cycle.

Especially important for understanding the place of the Emerging Technologies concept in the family of notions and relationships related to the innovation process is the understanding of the differences between the notions of incremental and radical innovations. The literature on technological innovation defines as "**incremental innovations**" the small improvements described by the learning curve, and by concepts such as "learning by doing," where improvements here are continuous, and where future improvements can be predicted causing relatively little disruption (Abernathy & Utterback, 1978; Freeman & Soete, 1997). On the other hand, "**radical innovations**" or "**technological discontinuities**" are the situations where a totally new technology comes along and displaces the incumbent technology. Examples are electric vehicles replacing gas-fueled vehicles, the USB flash drives replacing diskettes and CDs, etc. These changes are discontinuous, not continuous, and frequently cause significant disruption involving changes in industry leadership. Radical innovations frequently come from outside the mainstream, and are pioneered by small, entrepreneurial firms. According to Foster and Kaplan (Foster & Kaplan, 2001), nearly 70% of radical innovations come from outside the industry where they are used. Table 2 summarizes the main differences between the two types of innovation.

Table 2: Differences between incremental innovations and radical innovations

	Incremental	Radical
Technology	Innovation based on existing knowledge base, tools, processes, and technological paradigms	Innovation resulting from the exploration of new technologies
Trajectory	Linear and continuous	Sporadic and discontinuous
Uncertainty	Low	High
Idea Generation & Opportunity Recognition	Occur at the front end of the decision-making process; critical events are largely anticipated	Occur sporadically throughout the life cycle, often in response to discontinuities in the project trajectory
Process	Formal, slow and serial	Informal, flexible at early stages due to high uncertainties. Formal at later stages after uncertainties have been reduced
Strategic Factor	Improves competitiveness	Transform existing markets or create new ones

As Christensen (Christensen, 1997; Christensen & Raynor, 2003) points out, when a new technological innovation overturns the existing dominant technology in the market, it is called a "**disruptive technology**". This technology is both radically different from the leading technology and it often initially performs worse than the leading technology according to existing measures of performance. In contrast, "**sustaining technologies**" refers to the successive incremental improvements to performance that market incumbents incorporate into their existing product. According to the author, a disruptive technology comes to dominate an existing market by either[2]:

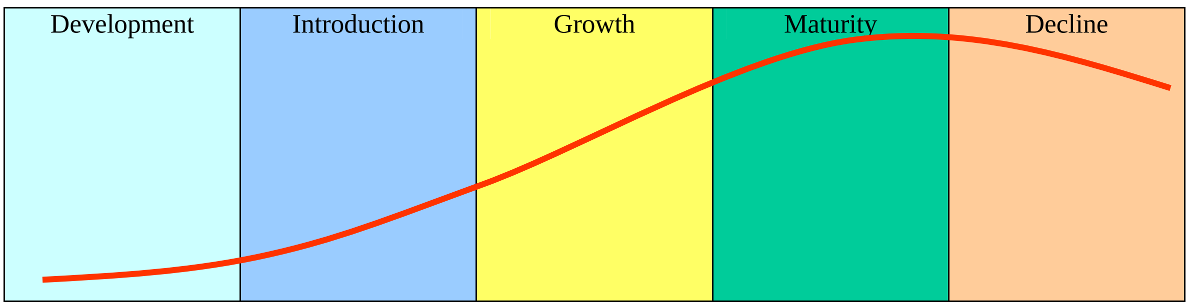
- Filling a technology gap that the older technology could not fill even if the disruptive technology is initially inefficient (as for instance more expensive, lower capacity but smaller-sized hard disks did for newly developed notebook computers in the 1980s), or

[2] Recently Christensen changed his notion of 'disruptive technology' for the notion of 'disruptive innovation' since, according to the author, what at the end is disruptive is not the technology itself but the innovation strategy instead. (Christensen & Raynor, 2003)

- Moving successively up-market through performance improvements until finally displacing the market incumbents (as digital photography has come to replace film photography).

An important characteristic of progress is that technologies evolve and reach limits (Freeman & Soete, 1997). A frequently seen pattern of evolution of a technology is the **"S" shaped curve**, where in the early stages the technology shows poor performance and the competitors usually underestimate the technology. In the middle portion of the curve, the technology takes off since some of the problems encountered in the first phase have been solved helped by learning curve effects, and customer acceptance has increased. In the final stage, no further improvement is possible. Figure 4 illustrates the cycle.

Figure 4: A Typical Technology Cycle



A typical "**technology cycle**" includes the following types of technologies:

- **Emerging Technology** - a technology that shows high potential but hasn't demonstrated its value or settled down into any kind of consensus. At this stage potentially successful technologies can be confused with sporadic "**hypes**". Hypes are unjustified, excessive discussion and usage of a technology or concept (hard to know beforehand). Hypes can be generated both by companies seeking financial investment or gain from their emerging technology, or academic researchers seeking notoriety for their research (i.e. ITs, Nanotechnology, Biotechnology, MEMS, Cold Fusion…). Two associated mistakes in the process of identification of Emerging Technologies is to focus on the first part of the growth curve, either assuming an eternal exponential growth, or becoming disappointed by the seemingly slow linear growth. In this stage, uncertainty and risk of failure are elevated, but the rewards seem very attractive.
- **Leading edge Technology** - a technology that has proven real potential in the marketplace but is still new enough that it may be difficult to market it.
- **Prevalent Technology** - when most customers agree that a particular technology is the right solution.
- **Dated Technology** - still useful, still sometimes implemented, but a replacement leading edge technology is readily available.
- **Obsolete Technology** - has been superseded by state-of-the-art technology, rarely implemented anymore.

Path-Dependence, Lock-ins, and Scientific Revolutions

Based on the current theories of technological progress, to assess the potential of an Emerging Technology, additional aspects are needed to consider. These refer to the nature of the technology and the probability for a new, disruptive innovation to emerge.

Most technologies are object of "**path dependence**" and "**lock-ins**". This happens when the outcome of a process depends on its past history, on the entire sequence of decisions made by agents and resulting outcomes, and not just on contemporary conditions (Islas, 1997; Leydesdorff, 2000). Bandwagon and network effects are at the origin of path-dependence. They lead to a reinforcing pattern, in which industries tip towards one or another product design. An example is the computer software market; each user is bound by the compatibility requirement.

Paradigm shifts and scientific revolutions are also relevant concepts to take into account in the conceptual map drawn here to facilitate the understanding of the context of the notion of Emerging Technology and the associated challenge of its measurement. According to Kuhn (Kuhn, 1966), a "**paradigm**" is a set of practices that define a scientific discipline during a particular period of time. It is not simply the current theory about the world around us, but the entire worldview in which it exists, and all of the implications that come with it. In this framework, scientists do what they think is worth doing and neglect potentially important gold mines. The science and the technology developed are thus bounded by the dominating but probably myopic consensus. A "**scientific revolution**" occurs as a result of a paradigm shift, that is, the process and result of a change in basic assumptions within the ruling theory of science. A paradigm is replaced when a) the demands for technological solutions to widespread problems are elevated, b) there are no solutions available to such problems, but c) some serious work has been done in this direction advancing progressively and experiencing positive feedbacks. As will be shown, these criteria will reveal relevant for the definition of Strategic Emerging Technologies.

Strategic Emerging Technologies

Considering the implicit strategic nature of the concept of Emerging Technologies and of its identification for the purpose of supporting the development of 'disruptive innovations' that aim to solve identifiable potential problems and can be supported by local efforts, the following definition could be useful to guide the operationalization of the concept and the design of the strategy to pursue. A **Strategic Emerging Technology** is one that exists in an area where a) a spurt of scientific activity has been identified; b) a sharply increasing level of technological activity is observed; c) there are identified unattended problems/needs (both current and future) and potential markets; d) local S&T potential to develop the technology has been identified ; e) there is identified open access for the Emerging Technology to flourish; and f) local comparative advantages exist.

Notice that this definition requires all the criteria mentioned to be met. The normative claim implicit in this definition is that the identification of Emerging Technologies not only needs to be done based on the information available but on the needs, capabilities, and potential for this technology to be successfully transformed into a disruptive innovation. This will require not only an exercise of data mining but also of needs- and capability- mining or identification.

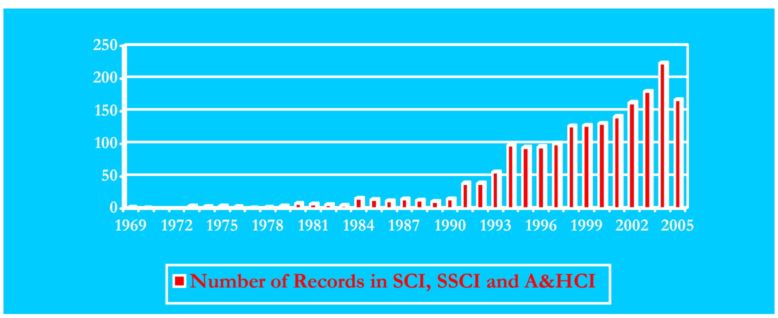
CHAPTER TWO: EMERGING TECHNOLOGIES AND RELATED TERMS

Definitions of Emerging Technologies

In the framework of this project, and after an extensive exploration of nearly 2000 articles in the literature of technology management, economics, and public policy, and other areas, we find that the concept of emerging technology has been defined and operationalized rather loosely since its first appearance in the 60s.[3]

As Figure 5 shows, the term emerging technologies, although an old one, is being used more and more frequently, especially since the 90s. It is used more frequently in the scientific journals than in the social science journals. We analyzed the data available using VantagePoint's natural language process and its statistical components as a tool, and found that the concept of emerging technologies is associated in the literature with several dimensions of meaning. Table 3 summarizes the ways the concept of emerging technology has been associated with a time dimension, a strategic dimension, a sectoral dimension, a role/type of technology dimension, and a disciplinary dimension.

Figure 5. Use of Term Emerging Technologies as Recorded in the Web of Science



[3] We conclude this based on the analysis of the abstracts of 1,927 documents registered by the ISI Web of Science published between 1956 and August 2005 containing the term "emerging technolog*", which included the articles published in a special issue of the journal Technological Forecasting and Social Change (Vol. 72, Issue 3) in 2005 that is devoted to the selected papers of the METiA (Managing Emerging Technologies in Asia) Conference. We found no definition of emerging technologies in any of the papers in the special issue. For an overall review of the METiA Conference and the special issue, see Koh, et al. (2005).

Table 3: Dimension of Use of the Term Emerging Technology
and Other Related Co-occurring Concepts

Dimension	Related concepts co-occurring most frequently with ET
Time-related	New Technology, Current Technology, Existing Technology, Technological Regime, Technology Development, Technological Advance, Technological Change, and Technology Maturation,
Strategic-related	Promising technology, high Technology, Economical Technology, Technological Innovation, Technology Adoption, Technology Diffusion, and Disruptive Technology.
Type/Role -related	Complementary Technology, Alternative Technology, and Applied Technology
Sectoral-related	Information Technology, Biotechnology, Nanotechnology, Health Technology, and Digital Technology.
Discipline-related	Management of Technology, Technology Forecasting/Foresight, and Economic Development.

A further analysis that included books, journals not indexed by ISI's Web of Science, and Google searches found that most materials used the term emerging technologies without defining it. However, a handful number of materials provide some operational definitions. Table 4 summarizes five of such definitions of emerging technologies.

Table 4: Definitions of Emerging Technologies

Source	**Definition**
(A.L. Porter, Roessner, Jin, & Newman, 2002)	"Emerging technologies are defined here as those that could exert much enhanced economic influence in the coming (roughly) 15-year horizon." 'Hot' R&D areas are those that show both strong recent (technology's class codes for which some 10%, or more, of the total articles occur in the most recent full year: 1999/all years considered*100>10), and increasing (growth rate between 1996 and 1999>2), research interest.
(Corrocher, Malerba, & Montobbio, 2003)	"Technology i is an emerging technology, if its growth rate in terms of patents between the period (95-96) and the period (98-99) is above the average of the sample which includes all the technologies"
(Hung & Chu, in press)	"Emerging technologies are the core technologies, which have not yet demonstrated potential for changing the basis of competition"
(Day & Schoemaker, 2000)	Definition 1: "Emerging technologies are science-based innovations that have the potential to create a new industry or transform an existing one."
(Day & Schoemaker, 2000)	Definition 2: "Emerging technologies are those where (1) the knowledge base is expanding, (2) the application to existing markets is undergoing innovation, or (3) new markets are being tapped or created."

In contrast to general definitions of emerging technologies above, Halal et al. (1998) (Halal, Kull, & Leffmann, 1998) has alleged to provide definitions of 85 emerging technologies in an appendix. However, it looks more like a list of specific technologies that the paper expects to emerge in the future instead of definitions of emerging technologies.

Since the handful of definitions of emerging technologies that our search has found are diverse, it seems that any agreement on the definition of emerging technologies has not yet been reached, though the term has been quite widely used in the literature. However diverse, the several attempts to define emerging technologies show that some authors have started pondering the meaning of the term. The concepts reflected in the definitions of emerging technologies can be summarized four-fold as follow: 1) fast recent growth; 2) in the process of transition and/or change to something new; 3) market or economic potential that is not exploited fully yet; 4) increasingly science-based. Table 5 below summarizes the afore-mentioned four major concepts as reflected in the definitions of emerging technologies.

Table 5. Major Concepts Reflected in Definitions of Emerging Technologies

Major Concepts of Emerging Technologies	**Sources of Definition**
Fast Recent Growth	Porter, et al. (2000); Corrocher, et al. (2003)
Transition / Change to Something New	Hung and Chu; Day and Schoemaker (2000) Definition 1 & 2
Market or Economic Potential	Porter, et al. (2000); Hung and Chu; Day and Schoemaker (2000) Definition 1 & 2
Increasing Science-based-ness	Day and Schoemaker (2000) Definition 1

The first concept of emerging technologies as incorporated in the definitions of emerging technologies in Porter, et al. (2000) and Corrocher, et al. (2003) is fast recent growth. Although the rate of growth might be more an indication of the importance of new or existing technologies, it can be a good proxy for the emergence of new technologies. Also, this concept was a key for those two studies in deciding upon the methods that were used to select specific emerging technologies. This topic will be covered further below when we discuss methods used for identifying emerging technologies.

Since the term emerging technologies connotes an evolution to something else, it is not surprising that some definitions incorporated the second concept of being in the process of transition or change to something new. It is notable that many definitions seem to relate this second concept of emerging technologies to the third concept keyed on market potential. Thus, the definition in Hung and Chu has referred to “changing the basis of competition” while Day and Schoemaker (2000) emphasize the creation of a new industry and transformation of an existing industry. Such a close relationship is not surprising because the commercialization process has been recognized to play a pivotal role in transitioning emerging technologies to new industries (Hung and Chu). However, the second concept based on transition and change incorporates non-market based meaning like expansion of the knowledge base as in Day and Schoemaker (2000). In turn, expansion of knowledge has been considered to include activities such as finding a new application of an existing technology or knowledge. In relation to the last point, we should note that the second concept of emerging technologies encompasses not only radical innovations but also incremental ones. As Day and Schoemaker (2000) note, emerging technologies include discontinuous technologies derived from radical innovation as well as more evolutionary technologies formed from previous research streams.

The third concept of emerging technologies is the most basic in a sense that most definitions reflect the concept of market potential. In relation to this concept, it is critical to choose a time horizon to consider in defining and/or selecting the emerging technologies as Porter, et al. (2000) has indicated. An additional point that deserves mention is that this third concept of market potential combines a potential to transform an existing industry/market with a potential to create a new one. This point is implied in the definitions of both Corrocher, et al. (2003) and Day and Schoemaker (2000). Thus, it can be said that an emerging technology can be an incremental one originating from its potential to change an existing industry, in addition to being a radical one originating from the potential to create a new industry.

The fourth concept of emerging technologies, the science base of the technology, is represented in the first definition of Day and Schoemaker (2000). This dimension should not be a surprise given the fact that developing many technologies having a revolutionary impact on the economy and society depend upon advances in science (Martin 1995). Also, it is in line with the recent trend that many innovative technological developments have been based on scientific advances; nanotechnology and biotechnology are good examples.

Other Candidate Terms for Emerging Technologies

Among the keywords that co-occur with the term emerging technology, disruptive technology is the only one that might serve as a candidate synonym for emerging technologies. In a similar situation, the concept of high technology has also served not only as a characteristic but also and sometimes as a synonym for the term emerging technology.

Disruptive Technologies

Chapter One already discussed the relatively new concept of distruptive technologies. This section explores whether or not disruptive technologies can be a good candidate concept for emerging technologies by comparing the definitions and concepts of both terms that might be close in meaning, judging by their frequent use together in the literature.

It seems that disruptive technologies has been defined in several different wasy, each of which has a different focus (Kassicieh, Kirchhoff, Walsh, & McWhorter, 2002; Ronald N. Kostoff, Boylan, & Simons, 2004).[4] There does not seem to be one widely accepted definition (Kostoff et al. 2004); indeed the various writers are still debating what definition might be appropriate (Walsh & Linton, 2000). Regardless of the state of this discussion, some definitions of disruptive technologies seem relevant or comparable to the aspects of emerging technologies as summarized in Table 6.

---

[4] Such definitions of disruptive technologies have been based on the following various focuses: firm-based product technology factor; industry wide product-technology factors; the gap between substitutable technological learning curves on cost or performance basis; customer behavior; product newness; market factors; and some combination of these factors.

Table 6. Definitions of Disruptive Technologies

Source	Definition of Disruptive Technologies
Kassicieh et al. (2002) Kostoff et al. (2004)	"Scientific discoveries that break through the usual product/technology capabilities and provide a basic for a new competitive paradigm"
Danneels (2004)	"A disruptive technology is a technology that changes the bases of competition by changing the performance metrics along which firms compete"

Before delving into the differences between the two terms, we note that there are several signs of a possible fit between the two. One is that some researchers have referred to disruptive technologies as emergent technologies (Danneels 2004; Kostoff et al. 2004). Additionally, certain definitions of disruptive technologies such as the ones in the above table share some major concepts reflected in the definitions of emerging technologies. Thus, the above definitions of disruptive technologies seem to incorporate at least the following main concepts of emerging technologies: transition/change, market potential, and science base. However, some concepts of emerging technologies such as fast recent growth are not shared by the definitions of disruptive technologies.

In addition, more emphasis is given to market and competition by the definitions of disruptive technologies than is true for emerging technologies. For example, when connoting a concept of transition and change, the definitions of disruptive technologies focus solely on market competition whereas the definitions of emerging technologies incorporate some non-market related concept like expansion of the knowledge base. Another distinction between the two terms can be drawn based on the fact that emerging technologies seem to encompass radical innovations as well as incremental innovations, with which disruptive technologies are generally considered to disassociate. Finally, we should note that there are various definitions of disruptive technologies with different focuses other than the ones that we selected as relevant to emerging technologies. Thus, those other definitions of disruptive technologies can diverge significantly from the concepts of emerging technologies. Based on the previous comparisons, it seems safe to conclude that the term disruptive technologies cannot be treated as equivalent to the term emerging technologies, although both terms seem to share some common concepts.

High Technology

The term high technology is also frequently used as having similar connotations as the notion of emerging technologies. In general, it is used to describe the firms, occupations, products, and activities associated with a set of industries designated as high technology. Like emerging technologies, high technology industries depend heavily on science and technology activities, including research and innovation; this dependence results in the frequent introduction of new or improved products and services. High technology firms generally have substantial economic impact, fueled by large research and development expenditure, capital investments, and a higher than industry average sales growth, employment of scientists, engineers and technicians, and average wages. The industries are often considered to have strong implications for productivity, international competitiveness, national competitiveness and the general standard of living (Riche, Hecker, & Burgan, 1983).

There is, however, no one established method for identifying which industries fall into the high technology category. Hence there is no consensus on which industries should be included in the group. Different scholars, organizations, regions and states apply different criteria to determine what falls into the category. Chapple et al. (2004) review recent definitions of high technology industries that include technology-producing and technology-using industries or both to gauge the degree of "high techness." As Chapple et al. note, typically the definitions result from considerable subjective judgment on the part of the researcher in establishing the bounds of the definition (Chapple, Markusen, Schrock, Yamamoto, & Yu, 2004).

The AeA (formerly the American Electronics Association) publishes a popular annual report, the most recent being Cyberstates 2005: A State-by-state Overview of the High Technology Industry. The report provides information on a number of indicators on investments, venture capital, employment, sales, and exports for high technology industries in U.S. states. The AeA uses a relatively narrow definition for high technology industries, which includes electronics manufacturing (computers and consumer electronics, communications, semi-conductors, components, devices etc, plus communications and software services. The AEA excludes biotechnology, pharmaceuticals and aerospace, and Chapple et al. argue that the bias reflects the focus of the membership of the organization, which is an electronics trade association.

In the 1999 Milken Institute study of America's High-Tech Economy: Growth Development and Risks for Metropolitan Areas, "high technology industries were defined as those with above-average expenditures on research and development and an above average share of technology using occupations – scientists, engineers, mathematicians, and programmers". Industries included were drugs, aircraft, guided missiles, and motion pictures.

The US Bureau of Labor Statistics has modified their definition of high technology industries progressively over the last decade. The definition is largely based on occupational criteria (Hecker, 1999; Luker & Lyons, 1997), with the most recent being that "industries are considered high tech if employment in both research and development and in all technology oriented occupations accounted for a proportion of employment that was at least twice the average for all industries in the Occupational Employment Statistics Survey" (Hecker, 1999).

Chapple et al (2004) also adopt a human capital (occupational) based definition of high technology industries, defined as those that have 9% or three times the national norm of S & T occupations or higher. They differ from the BLS in that they use a broader set of S&T occupations to define the industries and these are determined after careful consideration of the nature of the jobs and consultations with experts in science and engineering. The selected occupations include managers with science and engineering backgrounds, certain groups of computer professionals, petroleum and automotive engineers including designers.

The OECD classifies an industrial sector as high technology depending on the sum of direct and indirect R&D intensity[5]. The direct intensity is defined as the ratio of R&D expenditure to value added for each sector while indirect intensity takes into account the technology (R&D expenditure) embodied in intermediate and capital goods purchased on the domestic market or imported. The technical coefficients of manufacturing industries extracted from input-output matrices are used to calculate indirect intensity. The OECD definition covers only manufacturing industries and three groups of industries are defined high, medium (medium-high and medium-low) and low technology industries.

Emerging technologies, whether based on basic science, incremental or radical innovations and disruptive technologies can come from industries designated as high technology, but are not limited to only these industries. Incremental innovations are possible from so-called medium or low technology industries, but these are less likely to produce the more radical changes.

Relationship between Terms Potentially Related to Emerging Technologies

As this review and analysis of the literature demonstrates, there is a tendency for many authors, especially scientists, to use the term emerging technologies as a somewhat fancy term to mean new technologies. Additionally, our analyses of the phrases in abstracts indicated a possibility that some term such as promising technologies have carried meanings close to emerging technologies. Furthermore, we identified another term, disruptive technologies, to be related to emerging technologies. In practice, emerging technologies have also been associated with high technologies, in the sense of being science-based. Given the finding that several terms might be closely related with emerging technologies, this section concludes by exploring the relationships among those potentially substitute terms that some authors of the literature have been using somewhat confusingly.

Although some authors use new technologies and emerging technologies interchangeably, for most authors, the term emerging technologies means something more than just new technologies. In particular, several recent definitions of emerging technologies have incorporated some additional concepts other than just newness. Since not all "new technologies" are "emerging technologies," emerging technologies are a subset of new technologies. Although "promising technologies" have been used as a generic descriptive term for "emerging technologies,"every promising technology is also not an emerging technology. Thus, emerging technologies seem to be a subset of promising technologies, which are in turn a subset of new technologies.

[5] OECD Fact Book 2005: Economic, Environmental and Social Statistics http://miranda.sourceoecd.org/vl=8076588/cl=14/nw=1/rpsv/factbook/06-02-04.htm Accessed October 22, 2005

On the other hand, another related term, <u>disruptive technologies,</u> seems to share some characteristics with emerging technologies. However, they are not equivalent to each other. There are conceptual differences between the two, and the term <u>disruptive technologies</u> carries a variety of definitions with different foci.

Finally, the relationship between the concepts of high technology and emerging technology is again a fuzzy one. While in practice one can hardly imagine an emerging technology that is not a high technology, one cannot logically assume an automatic connection. In theory, emerging technologies can also be medium and low technologies as defined by current literature. Less problematic is the other way around, that is, we certainly cannot assume that the all high technologies are emerging technologies. It seems that the factor “time”— that is, newness – is what distinguishes one from the other, but even this assumption is not without question.

CHAPTER THREE: METHODS FOR IDENTIFYING EMERGING TECHNOLOGIES

As we have seen in Chapters One and Two, the concept of emerging technologies is widely used but seldom defined. Our literature review reveals that it is even less often operationalized quantitatively; that is, quantitative measures have very seldom been used to *identify* emerging technologies. Emerging technologies are in fact most often identified through qualitative processes involving expert judgment. Most of the quantitative studies of emerging technologies have taken the results of qualitative processes as a starting point, and characterized pre-identified areas quantitatively, a process we call *measurement* of emerging technologies. The measurement of emerging technological areas is much more common than the use of quantitative methods to identify emerging technologies.

This chapter puts both sets of methods into context by briefly describing some of the qualitative approaches. We then provide a generic description of the sources of data and approaches available to identify emerging technologies. The sources of data characterize different stages of the innovation process, and we use their placement in relation to the linear model to evaluate their potential for detecting emerging areas.

To live up to the conclusions of our conceptual discussion above, quantitative approaches to identifying emerging technologies face a triple challenge: identifying emergence, sorting technologies from sciences, and finding early indicators of market potential. The chapter discusses briefly the approaches theoretically available in literature-based analysis for identifying emerging technologies, evaluating them in relation to their effectiveness in meeting these three challenges. We conclude the chapter with examples of several quantitative approaches actually found in the literature. The next chapter goes on to describe how emerging technological areas can be characterized using VantagePoint.

Qualitative methods

Qualitative approaches to examine technology futures have evolved from the use of single methods and the anticipation of a single outcome to the adoption of multiple methods, criteria and the anticipation of many possible futures (Cuhls, 2003). Early approaches to technology forecasting and assessment such as scenario building, Delphi surveys, and expert panels are now part of more comprehensive methodologies such as foresight studies. Many countries, both industrialized and industrializing have now adopted foresight studies to inform strategic policy and planning activities in science and technology and other areas.

Delphi and Critical Technologies

Delphi surveys, which involve repeated surveys of individuals, usually experts in a particular field using the same questionnaire are useful for collecting and synthesizing information. They provide mechanisms for controlled communication and exchange of opinions without the dominance of a single individual. Delphi surveys make the tacit knowledge of experts about the future more explicit, and are useful for longer term assessments for which extrapolations make no sense. Since it is done anonymously, it can capture opinions on developments that might not have substantial evidence or support (FOREN, 2001). The Japanese used Delphi surveys to identify future developments of technology since the 1970s, but these are now part of more comprehensive technology foresight studies (Eto, 2003).

The U.S. government adopted the use of expert panels in its efforts to identify critical technologies in a series of biennial studies conducted over the period 1989-1999 (Wagner & Popper, 2003). The panel was charged to produce a list of 30 technologies that were essential to long term national security and economic progress of the US. Decisions were arrived at following deliberations by a 13-member panel of experts drawn from the public and private sectors. The panel received technical support in the form literature reviews on technological developments, background papers on specific technologies, and the organization and planning of meetings, initially from a private consulting firm then from the RAND Corporation. According to (Wagner & Popper, 2003), the limited use of the results emanating from the deliberations stemmed from a number of factors including ideologies, decentralization of the management of science and technology and the absence of mechanisms to link S&T policies with broader policies.

Foresight

The Foresight process seeks to bring awareness of long term social, economic and technological developments and needs to current decision-making (FOREN, 2001). It represents systematic and participatory efforts to gather intelligence on a wide of range of factors to inform medium-to-long- term vision building (Langenhove, 2002). A key aspect of foresight is its participatory element and efforts to build networks that involve key drivers of change, experts and other participants from the wider society. This serves to broaden perspectives, increase understanding and flexibility and reduce conflict in policy-making. Science and technology foresight has the potential to identify S & T related issues including emerging technologies and contribute to the strategic setting of research policies and priorities while taking into account the social aspects related to technology success (FOREN, 2001).

Foresight draws on widely distributed knowledge; it can be pursued using top down (more formal methods, e.g., Delphi) or bottom up (more participatory) approaches; it may be considered to be exploratory (looking from the present outwards to the future) or normative (adopting a future position and working backwards to identify how to get there); and in addition quantitative or qualitative methods or a combination may be adopted (FOREN, 2001). The end-result is expected to be a shared vision or plan with some levels of commitment from stakeholders because of their participation in its preparation. Foresight can be used as complement to other policy, planning or strategy building activities and is best done as a recurrent rather than a one time process.

The objectives of the foresight exercise should be clearly defined at the outset. The exercise typically involves the examination of time horizons of five to 20 years and may take six months or more to complete. Hence the process requires commitments of time from the participants, resources to cover costs, and sponsorship. Foresight exercises can be undertaken using a broad set of methods including desktop research, interactive brainstorming as well as broad participatory arrangements (Langenhove, 2002). Table 7 summarizes some of the methods used in foresight exercises.

Despite its potential and the undertaking of several technology foresight exercises in several countries in Europe, Asia and other parts of the world, several concerns remain about the effectiveness of the process and how it is conducted. The concerns, discussed in a forum on the *"Role of Foresight in Setting Research Priorities"* in 2002 provide insights on strengthening future foresight initiatives. Issues raised about foresight include the fragmented nature of national and European exercises, with few links between the foresight activities or other policy initiatives and much duplication of effort (Busquin, 2002; Langenhove, 2002). Up to 2001, 26 technology foresights were undertaken in Europe in eight key sectors of the economy including chemical, energy, and the environment. In many foresight exercises, there is ambiguity in the objectives, whether the focus is on product (key technologies) or process (participation and networking). The type of participation in the process was also questioned as it was felt that there was not enough involvement of the policymakers and participants were not chosen in a transparent manner (Haugg, 2002).

In addition there is concern that foresight exercises were not being evaluated in areas such as the translation into policy, the extent to which it increases awareness, its contribution towards building networks and lessons learned from successes and failures (Georghiou, 2002). Other issues stem from difficulties associated with consensus building and the loss of "out of the box" ideas, and the possible hijacking of the process by those with vested interests. In addition, the report points to the need to take into account the possible risk of inefficient domestic R & D investment (including highly qualified human resources) given international specialization in S & T areas and increasingly globalized technology markets; the increasingly multi-disciplinary and cross-sectoral nature of major technological breakthroughs as well as the importance of serendipity in such breakthroughs; and the roles of public and private investments in R&D (Malkin, 2002).

Keenan describes the efforts of the UK to identify and prioritize emerging generic technologies in a foresight exercise undertaken in the period 1994-5 (Keenan, 2003). The foresight exercise included the typical elements, a client a steering group, a set of thematic panels, external participants and support staff including knowledgeable foresight practitioners. Delphi surveys were used to get wider views of the future, and this was followed by the use of expert panels to prioritize technologies. A generic technology was considered to be one that exploitation yields benefits for a wide range of sectors of the economy and/or society. The paper offers one definition of an emerging technology, attributed to the U.S. Department of Commerce that is "one in which the research has progressed far enough to indicate a high probability of technical success for new products and applications that might have substantial markets within approximately 10 years." (Keenan, 2003).

A concern in the exercise was the level of disaggregation or “granularity” of the technology to be considered. Too broad a category would reduce the level of focused action in a specific area, while more detailed categories produced a large number of areas that were difficult to appraise. In the setting of priorities, both science-push and demand-pull were considered. Although a prioritization model was developed as part of the exercise, several modifications were made during the process in order to arrive at the final list of priorities (Keenan, 2003).

Table 7. Broad classes of Foresight Methods

Criteria	Methods
1. Methods based on eliciting expert knowledge to develop long term strategies	– Delphi method – Expert panels – Brainstorming – Mindmapping – Scenario analysis workshops – SWOT analysis
2. Qualitative methods that make use of statistics and other data	– Trend extrapolation – Simulation modeling – Cross impact analysis – System dynamics
3. Methods to identify key points of action to determine planning strategies	– Critical /key technologies – Relevance trees – Morphological analysis

Source: Foresight for Regional Development Network: A Practical Guide to Regional Foresight
ftp://ftp.jrc.es/pub/EURdoc/eur20128en.pdf

Data sources

Information on scientific research is organized at different levels for example by disciplines, fields, specialties or research domains in unstructured (full text documents, web pages) or structured formats (awards, publications or patent databases). In the analysis of emerging technologies, data bases can be selected depending on the objectives, information available, and costs. In evaluating the suitability of data sources, considerations include coverage, biases, content quality, record structure, and keyword availability. Databases treat keywords differently and vary on the consistency they maintain with keyword structure. For example, *MEDLINE* uses MESH or hierarchical structure and is very consistent, while the Web of Science is less consistent and *Pascal* is not consistently structured.

Databases on awards for specific research activities such as the *RADIUS* database and those from the individual agencies for example, the National Science Foundations, provide information on research funded by the U.S. Federal government. Because the information is gathered from the proposals that were submitted and is therefore forward-looking, it can provide early indications of significant or growing activity in a particular research area. Although the *RADIUS* database is a fairly comprehensive source, it is not complete as not all awards may be reported. In addition, different agencies use different reporting formats.

The *Science Citation Index* and the associated indexes made available through the Web of Science are leading sources of information on publications in a broad range of disciplinary areas. *MEDLINE* provides a comprehensive source for medical and biosciences research, *AGRICOLA* for agricultural research, *Chemical Abstracts* for chemistry, and *INSPEC* and *Compendex* for engineering research. Only the Web of Science data bases provide citation information, and they also include the most comprehensive address information. Their weakness is the lack of systematically applied indexing terms.

Several patent databases provide information on the subset of inventions and innovations that need to be protected with patents. The most comprehensive source is the data provided by the U.S. Patent and Trademark Office (USPTO) because most companies want to protect their intellectual property in the large U.S. market. The analytic versions of the U.S. patent data include citations, both to other patents and to the scientific literature; these can form the basis for some of the analytic techniques discussed in the next section. The European Patent Office and World Patent Office also have data available for analysis. Because patents by definition lag the invention of specific techniques, they are unlikely to be particularly useful for identifying emerging technologies, although they may provide useful if lagging information to help with indicators of the commercial potential of emerging areas that are being tracked with other methods.

Business related sources on market research, venture capital and business start-ups also provide potentially useful information on downstream activities and trends. Even if the focus is primarily on the basic research, analyses of downstream activities can provide useful contextual information which can be used to complement publication or patent analyses. But data sources that are most closely tied to technologies and markets are lagging indicators that are unlikely to detect emergence early enough to allow effective strategic response.

Typically with publication databases, a time lag exists between the time that the research is done, publication and the time that it gets into the database. Frontier work (analyses of awards databases, websites of the research groups, or direct contact with research groups) could complement publication and patent analyses. Both upstream (R&D funding, personnel, projects, gray literature, confidential information sources) and downstream (businesses interests, popular attention, and relevant policy activities) activities can be examined in the analyses of emerging technologies. Analyses of emerging technologies, however, cannot be limited to the information available in formal databases. The effort also needs to include more qualitative information on social trends and tacit knowledge.

Quantitative techniques

As mentioned in the introduction to this chapter, analytic techniques for identifying emerging technologies face a triple challenge.

- First, they must be able to identify emergence, through a combination of detecting newness and measuring rapid growth.
- Second, they must be able to distinguish technologies from sciences. As our discussion of data sources just indicated, the earliest indication of new high technology capabilities is likely to appear in the scientific literature, which is also quite amenable to quantitative analysis. However, some of the areas that emerge in the scientific literature will obviously be the beginnings of new sciences, not new technologies. How can a method distinguish between the two?
- Third, once technologies have been identified, our concept of emerging technologies includes some indication of market potential. Is it possible to find such an indication in the research literature?

To meet these challenges, there are a finite set of analytic approaches to data sources like the ones we have just described. The analyst can use existing categories or indexing terms, or use data mining techniques to discover the structures inherent in the data. For the latter, one can use either citation-based or text-based analysis.

Indexing Categories and Vocabularies

The first option is to measure growth and characterize growing areas by using existing categories in indexing schemes or terms in controlled indexing vocabularies. This approach has both advantages and limitations.

- The primary indicator of newness is measured with the appearance of a new category, so the analyst is completely dependent on the indexer to identify emerging areas. The indexer is exposed to the natural language used in published articles, and so may be the best source of qualitative judgment about new concepts emerging in the literature. Nonetheless, inertia will work against the introduction of new terms.
- Another possible indicator of emergence within a controlled indexing vocabulary would be rapid growth in a particular pre-existing category. In several of our TPAC case studies, the period of growth that is generally identified as the "emerging technology" phase is characterized by a surge of growth within an existing category, rather than the appearance of a whole new category.

Can controlled indexing categories and vocabularies meet the second challenge, distinguishing technologies from sciences? Perhaps, at least in an approximate way. Most controlled vocabularies code technique and method separately from subject and even from experimental models. We hypothesize that this area of the vocabulary would be the richest source of indicators of the kinds of emerging capabilities that develop into technologies.

On the third challenge, existing controlled vocabularies might be mined for techniques that are seen as having market potential by using some data elements associated with articles within each category. First and foremost, the appearance of industrial authors could be an indicator. Additionally, abstracts and full text might carry terms like “industrial application,” “industrial use,” or “market potential.” Occasionally, a controlled vocabulary itself might also include categories that reflect industrial interest.

Data Mining

The alternative to direct use of indexing categories to measure emergence is to use more sophisticated information tools to find the structures inherent in the data itself. These techniques use the data elements associated with papers or patents to identify structures inductively. These structures become the units for further analysis. In principle, any text element that can be shared between two items in a data set can be used to find inherent structures. Three types are common in scientometrics: co-authorship analysis, co-word analysis, and citation analysis.

- Co-authorship analysis identifies social structures within data sets by using co-authorship as a similarity measure.
- Co-word analysis identifies intellectual structures using the co-occurrence of words or phrases associated with documents.
- Co-citation analysis identifies intellectual structures using the links created when two older documents are cited together in newer documents.[6]

Clustering and factor analysis are the most common ways to structure the data. In either approach, the stronger links between items are used to create the structures (either a cluster or a factor), and weaker links can be used to indicate links between the structures. Using the weaker links, maps can be created of the social or intellectual space through multidimensional scaling routines. We will refer to each cluster or factor below as a “group”

In order to detect emergence, this kind of analysis must be applied to a series of time slices, for example, a series of years of data (2000, 2001, 2002), or alternatively, a series of groups of years (e.g., 1999-2001, 2002- 2004, etc.) A group that appears for the first time in the map is *new*. If it appears again in the next slice, and if the number of associated papers or patents is going up faster than the average, the group may be said to be *emerging*. In order to trace such groups over time, a measure of similarity between one year’s areas and the next’s must be calculated. Some carryover information is needed. In co-authorship analysis, the authors themselves can carry over. In co-word analysis, terms can carry over. In co-citation analysis, cited documents carry over.

[6] A co-citation link is formed between two older documents when they are cited together in a newer document. Citation information can also be used in the form of bibliographic coupling, the link formed between two newer documents when the cite the same older document. Bibliographic coupling has not been applied to evolutionary analysis of the kind we are discussing here, so we will not refer to it further.

When items grouped together in one year appear in two structures in the next year, the area can be described as splitting. When items that appeared in separate structures one year appear together in the next, it can be said to be merging. The application of these tools to a series of time slices thus constitutes an analysis of the evolution of group structure. The challenge in identifying emergence is elegantly solved through such an approach. The ability of the approach to address the second and third challenges depends, as in the case of the indexing categories, on whether data elements are available in the underlying data set that provide indicators of technology status or market potential.

The preceding discussion is quite generic, and neglects two important practical issues in applying these techniques: the size of data sets and the specificity of the items grouped. The co-citation techniques (discussed in the last section of the chapter) have been developed to be applied to very large data sets. To accomplish this, the approach uses single-link clustering to group items, a technique that runs sequentially through a list of links rather than operating on a co-occurrence matrix. The disadvantage of single-link clustering is that it sometimes creates heterogeneous groups through chaining (grouping together several dense sets of relationships through a single link). Techniques that find structure by analyzing co-occurrence matrixes will have lower limitations in size than single-link clustering.

The second practical challenge is the specificity of items grouped. Authors have a reasonably limited range of people they co-author with; even the most coauthoring scientist will not be linked through co-authorship to thousands of others, and most will co-author with a dozen or fewer. Cited documents likewise are fairly specific to particular research areas, with the exception of widely cited methods papers. Keywords and natural language phrases, however, tend more often to be widely applied. Co-word approaches to evolutionary analysis must find ways to sort out the terms that are too general to be useful in finding specific areas. Co-citation analysis attempts this task through normalization of link strength.[7] Perhaps co-word approaches can adopt a similar approach.

Combinations

Finally, we should note that the two general approaches we have just described can be combined. An analysis can start with a data set retrieved through the tools of an established indexing scheme or controlled vocabulary, then analyze the internal evolution of that area using one of the data mining tools. The internal evolutionary analysis may then be combined with other ways of characterizing the area such as lists of the most prolific authors and institutions. VantagePoint has been designed to make these tasks easy, as we describe in the next chapter.

---

[7] The formula used is generally intersect over union, that is, co-citations over cites to item one plus cites to item two minus co-citations.

Three examples

Predicting emerging technologies with the aid of text-based data mining (Smallheiser, 2001)

Smallheiser uses text-based mining to predict genetic engineering technology that might impact viral warfare in the future (Smalheiser, 2001). The analyses were carried out using conventional *MEDLINE* searches and a package of advanced informatics techniques known as Arrowsmith. The approach combines detailed customized search strategies with expert analysis of results.

Smallheiser identifies "the critical factor as the overall strategy of approaching the problem: first, to define two specific fields explicitly, genetic engineering and viral warfare that are expected to have complementary information; second to identify common factors that bridge the two disciplines (i.e., research on viruses) and third to progressively shape the query once the initial findings are obtained" (Smalheiser, 2001).

In this paper, Smallheiser adopts a highly targeted or micro approach to identify intersections between two specific fields. This is in contrast to a macro approach that attempts to identify relationships from the examination of a large body of literature covering several disciplines. A specific research question is posed; the analyses focus on strong and consistent research findings that are reflected in the titles of papers although there is some perusal of abstracts and texts; and the initial search is limited to subsets of studies on genetic engineering related to the virulence of viruses and aerosol dispersion of viruses, key issues related to viral warfare. The initial search is broadened subsequently, and following the examination of the documents returned in the search, three specialized keywords, technologies for packaging viruses are identified and used as query terms for further searches. These are DEAE-dextran, liposomes and cyclodextrins. Candidate technologies are assessed based on the detailed examination of the papers retrieved.

The emergence of new technologies in the ICT field (Carrocher, 2003)

Carrocher et al. examine patent data to identify emerging technologies in the information, communications and technology field (Carrocher, Malerba, & Montobbio, 2003). Patent abstracts are selected because they provide a comprehensive description of the technology, product or process being patented as well as of the potential applications of the technology and this facilitates the identification of links between different knowledge domains and technological fields. They evaluate patent growth rates between two periods, 1995-96 and 1998-99, and identify emerging technologies as those that have a growth rate above the average of the sample that includes all the technologies. This technique thus exemplifies our first category, using existing indexing categories.

The authors choose patent abstracts from eight technological classes belonging to the sections of *Physics (G)* and *Electricity (H)* of the International Patent Classification from the EPO-CESPRI dataset for the time period 1995-1999. These are:

G01 – measuring and testing
G06 – computing, calculating, counting
G09 – educating, cryptography, display, advertising, seals
G11 – information storage
H01 – basic electronic elements
H03 – basic electronic circuitry
H04 – electronic communication technique
H05 – electric techniques not otherwise provided for

An ad hoc algorithm is used to extract sequential triples of words from a sample consisting of 102,547 patent abstracts. The sets of words or triples that appear with significant frequency are examined further to determine the technology, product or technological application that it represents. Methodological issues associated with the selection of keyword triples include the concealment of critical words by the drafters of the patent; linguistic patterns that are country specific; generic keyword associations and meaningless combinations of triples that have to be cleaned from the data, among others. Co-word analysis is used to assist in identifying meaningful combinations, and 119 triples that represent relevant technologies, applications, platforms, or products were identified for the analysis.

Carrocher et al. further examine relationships between patent applications, technological classes, firms, and countries that the triples identified as technologies are associated with. Triples or technologies vary in the level of distribution in technological classes, with some triples appearing in several technological classes. Some triples closely overlap a technological class and these are often associated with a specific product such as “cathode ray tube” or “lithium secondary battery.” Other triples are distributed across several technological classes and tend to be associated with technological platforms, e.g. “asynchronous transfer mode” or “graphical user interface”.

A technology is considered emerging if the growth rate in terms of patents between the period 1995-96 and the period 1998-99 is above the average of the sample that includes all the technologies. From the examination of Herfindahl indices, which provide measures of the concentration of an activity in a group compared to the whole, they find consistent with theory, that the inventions and patent activities related to emerging technologies are relatively more concentrated across countries, firms and specific technological classes. In non-emerging technologies, there is greater hybridization of technologies. Patterns of knowledge flows which are assessed using an index of originality are also different for emerging and non-emerging technologies. According to the findings of Carrocher et al, knowledge for emerging technologies seems to be concentrated in the same technological class, which is interpreted to mean that ETs relate to a narrower technological domain and rely on more specific sources of technological knowledge. For emerging technologies, although innovative activity is concentrated in a few firms, the firms are more likely to draw on knowledge from internal and external sources while firms involved in non-emerging technologies draw on less dispersed knowledge sources.

Visualizing science by citation mapping (Small, 1999)

Co-citation analysis of the scientific literature has been applied most extensively to evolutionary analysis, in the work of Henry Small and his associates at the Institute for Scientific Information. A strength of Small's approach is that it was designed to be applied to very large data sets, and can thus be used to search broadly for emerging areas on an interdisciplinary basis across a broad range of scientific activity. The measures of emergence in his system are clear-cut. It thus meets the first challenge easily, and in fact is the leading published technique for scientometrics evolutionary analysis.

The power of the approach to find inherent structure in large data sets can be illustrated with a specific example from Small's work. According to Small, the spatial representation of scientific literature can facilitate the understanding of conceptual relationships and developments. Small argues that maps of science produced from citation analysis reveal changes in the structure of science and the state of knowledge over time. Specific discoveries, methods, or ideas that are shared by authors become apparent when the citations in highly cited documents are examined (Small, 2003). In (Small, 1999), a simplified method of ordination and hierarchical nesting is used to produce maps of science from citation analyses. The maps can be analyzed subsequently to identify cluster linkages and to discover new pathways through science.

In order to select cited articles and take into account variations of citation and reference intensity, Small establishes an initial threshold (papers cited 5 or more times) of papers to select; then uses fractional citation count, an inverse weighting based on the total number of articles with 5 or more cites in the reference list of the citing article, to further prune the articles. Cited documents are used if the sum of the reciprocal of the weights derived from the citing articles is greater than one, yielding a final sample of approximately 164, 000 cited documents. The highly cited papers are used as markers for individual topics in the representation of scientific structure.

The selected documents were partitioned in to 5-level hierarchy of clusters using the co-citation linkage method, followed by ordination of objects within the cluster then the integration of local structures. Small uses the science mapping program, SCI-Map which arranges documents in two dimensions using geometric triangulation to produce two dimensional co- citation maps. At the first level of aggregation, the clusters consist of documents, which are visually represented as document circles are determined by their citation frequency. Subsequent levels enclose lower levels and at the second and higher levels, the size of the circle or object is dependent on several factors including the number and sizes of lower-level objects contained. The radii of circles in the map are measured in Garfield units. The labels of the objects are based on a frequency analysis of article titles and journal category names and lines in the map represent strong co-citation links among clusters.

The results detailed in (Small, 1999) show four main topic regions, physical sciences, a biology region, medical sciences and a behavioral /social sciences region that characterize the 35 level-4 clusters representing the major disciplines. The exploration of subject area takes place by drilling down from the broader disciplinary level and progressively focusing on the document level. Small identifies “hot fields” as clusters with a high number of recent papers and a high mean year of publication. He demonstrates the progressive analysis of the biology region and sub-regions showing a level-3 cancer genetics cluster, and eventually levels-2 and 1, where the frequency and age of particular documents can be compared.

The initial analyses found that the maps produced were skewed towards the physical sciences and away from the biological sciences, which Small attributes to the threshold used in fractional citation counts. These can be adjusted to produce a more balanced representation of fields. In addition other strategies can be adopted for example using more hierarchical levels, combining other indirect citation link forms with co-citation In concluding, Small argues that the method is a useful heuristic device to visibly organize information and the linkage patterns produced offer the opportunity to explore extended knowledge pathways.

CHAPTER FOUR: ANALYSES OF EMERGING TECHNOLOGIES USING VANTAGEPOINT

"Tech mining" of science and technology information has the potential to enhance national research and development investment strategies by profiling internal and external activity in target research domains or areas of strength; detecting new and potentially disruptive technologies; and assessing the R & D program. For R & D programs, the process can yield benchmarks to identify national strengths or gaps, institutional strengths, potential areas of collaboration and roadmaps of the best prospects to attain goals. Key process considerations include the identification of target users and their competitiveness technology investment needs; resources to be used; and potential technology intelligence products (analytic products, scientific breakthroughs, emerging trends). The technology intelligence products are reviewed and validated subsequently by experts, who remain a vital part of the process. The entire process is considered to be iterative as the feedback serves to refine the technology intelligence products and the analytic processes that provide the information for decision-making.

The analyses can be approached using either extrapolative or normative frameworks. In the extrapolative approach, the data is accepted or taken as given and trends are extrapolated. This approach provides descriptive or benchmarking information, which is interesting but does yield appreciable insights. The normative approach looks to the desired goal for the future and how to achieve this therefore it is action oriented and addresses information needs of target users.

Information is collected on research and commercial activities such as funding awards, publications, patents and business related activities (venture funding, business start-ups or market research). These include details about the initiators, institutions, innovators and ideas. The *VantagePoint* software facilitates the tech mining process by increasing the efficiency of analyses a large number data and the subsequent presentation of information. More detailed elaborations of the use of *VantagePoint* can be found in (A. L. Porter & Cunningham, 2005); *VantagePoint* Help Manual, in papers at http://www.tpac.gatech.edu/papers.php and the additional references listed.

"Tech Mining" Methodology

VantagePoint is software used to discover knowledge in text databases by applying data and text mining techniques to identify patterns and provide different perspectives on large collections of text. The patterns are identified through the co-occurrence of words (terms) in structured information sources. Co-word relationships do not generally make sense across a set of unstructured documents, therefore these are best analyzed using trained agents. The trained agent can structure data, which is fed afterwards into *VantagePoint.* Up to this point, no explicit algorithm has been established using *VantagePoint* to identify "emerging technologies." However, several bibliometric explorations have been consolidated to provide over 200 indicators that can be used to guide the identification of emerging technologies. The preceding sections discuss the conceptual landscape of “emerging technologies” which, in this section, includes (but is not limited to) the emergence of new scientific areas.

The process to identify emerging technologies begins with the careful consideration of objectives, the time frame for which the analyses is being done (e.g. long - 15 years or short – 2 years), and the decisions that need to be made. Detailed management of technology questions are identified from the issues. Explicit empirical innovation indicators are then enumerated to address those questions. Table 8 provides a summary of nine technology management issues and examples of possible questions are outlined in Table 9. Details on 39 technology management questions and candidate indicators for each are provided in Table 13.2 of Porter and Cunninham (2005)(A. L. Porter & Cunningham, 2005). In addition, consideration should be given to the frequency in which the analyses are repeated. Monitoring of target technologies, related technologies and contextual factors is essential for understanding what is happening with the technologies and for effective forecasting (Watts & Porter, 1997).

Table 8. Technology Management Issues

R&D Project Initiation	**Collaboration in Technology Development**
Engineering Project Initiation	**Identifying & Addressing Competitive Threats**
R&D Portfolio Selection	**Tracking & Forecasting Emerging Technologies**
New Product Development	**Identifying & Assessing Breakthrough Technologies**
New Technology Development	

Table 9. Technology Management Questions

Example “What” Questions

- **What emerging technologies merit our ongoing attention?**
- **What facets of this technology development are especially hot?**
- **What are the component technologies that contribute importantly? Significant subtypes of the technology?**
- **How does this technological development fit within the technological landscape?**
- **What is driving this technology development?**
- **What are the key competing technologies?**
- **What are the likely development pathways for this technology?**
- **Assess the maturation of the component technologies?**

Example “Who” Questions

- **Who are the available experts?**
- **Which universities or research labs lead in this technology – overall or in particular aspects?**
- **Which companies lead in particular aspects (main topics) of this technology?**
- **How strong are the leading companies’ R&D teams?**
- **How do leading companies’ development emphases compare to ours?**
- **What other technological strengths does each leading company have?**
- **What smaller companies or individuals have attractive IP relating to this technology?**
- **Who’s partnering with whom?**

Tech Mining is considered more useful for analyses covering a medium time horizon of 5-10 years because quantitative methods are less useful for longer time frames (more than 20 years). *VantagePoint* analyses have the potential to identify trends of increasing concentrations of scientific and technological research activity over time. Note that the trajectory of scientific advances with innovation potential over the long term is much harder to predict than technological ones, as these are more likely to involve discontinuous advances or radical changes.

The Tech Mining activities are envisaged to take place in three phases. These are the:

1. Intelligence phase, involving planning for and collecting the data to be mined
2. Design and analysis phase, which consists of deductive and inductive analyses to derive knowledge from data
3. Choice phase, in which options are identified via Tech Mining and the appropriate selections made.

VantagePoint is a tool that helps the analyst; it does not replace him or her. As (Roy, Gevry, & Potenger, 2002) note, interpretations remain heavily dependent on the data and the expertise of the analyst. In addition, the results of quantitative or bibliometric analyses are viewed as complementary to information obtained from expert opinion and other technological forecasting methods.

Intelligence phase

The search process is critical for recovering relevant and potentially useful information and eliminating irrelevant ones that can result in the identification misleading trends. In the intelligence phase, the following activities are undertaken:

- identify the right sources (where to search)
- formulate queries (how to search)
- download data (search results)
- survey and clean collected data
- summarize and describe data (use lists or tables)

Data Sources

Analyses with *VantagePoint* favor structured data sources, with the data records systematically organized in machine readable form (fielded, delimited). Appendix 1 lists a number of databases commonly used in *VantagePoint* analyses.

The data sources discussed in Chapter 2 are heavily biased towards English language sources. However these are the databases that TPAC analyzes most frequently using *VantagePoint*. Analyses of databases from other geographic areas such as the European Union, Japan, China, India and other ASEAN countries can be considered for a more complete coverage of emerging technology issues. A version of *VantagePoint* called TechOASIS is for US Government use. It developed a capability in conjunction with an automatic translation software, SYSTRAN, to handle non-Roman character languages. This was used by the Office of Naval Research's field office to assess Japanese language S&T information. *VantagePoint* has also been applied by Alisa Kongthon to examine Thai research records in the native language (without translation). The software matches character strings, so is able to do its basic and co-occurrence analyses in any language. However, the natural language processing (NLP) capability to parse sentences is attuned to English.

Query Formulation

The formulation of queries has to take into consideration the issue of scientific specialization, as well as recognize that many technological advances are increasingly science-based. In addition, when the emergence of new technologies is viewed as an evolutionary process of technical, institutional and social change, technology fusion, marked by the movement and recombination of technologies across systems and industries plays an important role (Corrocher et al., 2003). Thus the search query serves to link complementary sources of information and identify common factors that bridge the gap between different disciplines or technological systems (Smalheiser, 2001). Given that the potential of scientific discoveries in one discipline to impact another may not be readily apparent or recognized, the formulation of queries requires the expenditure of both time and effort.

The formulation of queries is an iterative process, in which the search and analysis steps may have to be repeated several times in order to achieve the desired results. The queries are formed following careful consideration of the objectives and can be based on key terms or phrases (descriptors or index terms), specific persons, places or articles. These can be combined using Boolean operators or winnowed using special features that may exist within a database. The objectives will determine the breadth (highly inclusive with few relevant items missed) or precision (narrow focus) of the search. Often searches using specific topics of investigation (terms in the field) yield more fruitful results when compared to searches done using the broader disciplinary titles or field names, similarly combinations of natural and scientific languages have the potential to yield richer results than either alone. Although much of the attention goes to the core R & D activities of the target arena, it is worth reaching out to novel ideas in other research areas i.e. the fringe or potential contributors. First there is need to get the core right; in an example taken from Kostoff on "water purification", the initial search is followed up with a search of purification of other liquids and with terms similar to purification – disinfection, separation, extraction (R. N. Kostoff, 2004)

Example: Query formulation

An initial search of the Web of Science to identify publications by academic and industrial researchers on "Nano chemical patterning" using the keywords chemical patterning, molecular patterning, chemical nanopatterning, chemical nanofabrication and molecular rulers yielded 903 records. The most commonly used keywords from the records were identified and passed to the "expert" in the field for review. The search strategy had returned considerable noise with off target keywords in the domain of immunology and genetics. After reading some of the abstracts it became clear that "chemical pattern" and "molecular pattern" were picking up papers about pattern recognition that was unrelated to nano-technology. The search strategy was revised to search for only chemical and molecular patterning and not its variants. In addition, other search terms, self-assembled monolayers, dip-pen nanolithography, nanocontact printing, and microcontact printing were included. The revised strategy returned 1074 records which are confirmed to be on-target.

In order to assess the quality of a search, multiple and redundant search terms can be included in the query and the search results compared to that of a more narrowly tailored search. This will yield a wider range of articles, but more irrelevant articles are likely to be included. Technically speaking, this would represent excellent "recall," but weak "precision." Perusal of the abstract or articles, and consultation with the experts in the field on the search terms and the articles collected, can also be used to assess the quality of the search. The consultations will yield additional search terms that may be useful to include in the search query as well as identify strategies to eliminate "noise" or irrelevant articles. Search queries should be saved so that they can be repeated or modified, depending on one's needs.

The selected databases can be searched individually or simultaneously and individual searches may be subsequently combined. Since typically large quantities of information are retrieved in text mining, one has to be cognizant of the limitations or restrictions that exist on information retrieval or subsequent analyses in different databases.

Data Collection and Cleaning

Depending on the database and the needs, the appropriate fields are selected from the database in order to answer questions on "who", "what", "where", or "when," as well as different combinations of these interrogatory terms. Potentially minable data elements include topics, authors, sources, citations, institutions, locations, and years. The use of existing categories in the controlled indexing vocabulary has the advantage of being easy to use but it it is unlikely to track emergence well. For example answering the question, which *MEDLINE* category is growing the fastest?, may say more about the indexing than about research activity patterns.

Example: Field Selection

In a Web of Science search on "biotech enzymes," the following fields were selected: author, author affiliation, country, publication type, journal, publication year, title, title NLP phrases, keywords (author) and keywords plus. For a similar search in the Derwent World Patent Index, the fields selected were: Derwent classification, family member countries, international classifications (Main) 4-digit, inventors, patent assignees, priority countries, priority years, title, and title NLP phrases.

Typically, the information from the database is extracted as text and converted to data on import into *VantagePoint*, using the appropriate filters or configuration files. *VantagePoint's* Import engine editor enables advanced refinement to capture particular facets in the data. The process simultaneously organizes the fields in the database into lists that can be used for the description and analyses of the data. Alternatively, the information may be exported directly to *VantagePoint* using 'ris' format, if available in the database.

Data cleaning is done by manipulating the thesaurus, fuzzy editor files, data fusion, list comparisons etc. to remove duplications, misspellings, hyphenation, capitalizations or stylistic differences in terms that can distort the analyses. Fuzzy matching is used to consolidate records with identical fields, variations on author names and duplicate records. Thesauri collect variations of the same entity so that consolidation of the various forms takes place. Specialized thesauri can be built and modified by the user to reflect interest area and needs. In addition scripts or macros based on Visual Basic or Java Script can be developed to automate data cleaning (and various analyses and information representations). However, the cleaned data should always be checked against the raw data to ensure that the desired results are obtained.

Data Summary and Description

Several options exist in *VantagePoint* for both describing and producing broad summaries of the data. These can provide ideas to expand the search; identify data problems, such as incongruent items; and point to additional directions for analyses. Summaries or basic analyses include lists, groups and maps.

- Lists show all the items in a particular field across all the records in the imported set.
- Groups are created from subsets of records.
- Breakouts are formed from the co-occurrence matrix of a subset significant field and another field of interest
- Maps are the visualization of a multi-dimensional co-word analysis that shows the relationships among chosen fields of data. Generally meaningful groups are

identified to reduce the number of terms so that computing resources are not tied up. Alternatively, multiple pages of maps may be requested with large datasets.

Several types of maps are available in *VantagePoint* including autocorrelation, cross-correlation, factor maps and principal components decomposition (PCD) maps. The auto-correlation map shows how selected items from one list relate to each other. These are often used to depict knowledge networks -- that is, patterns of co-publication among authors. The cross-correlation map shows how selected items from one list relate to items from another list. These are often used to visualize differences between organizations and researchers. For instance, the same set of authors can be mapped to show which ones use similar terminology. Comparing this map to the corresponding autocorrelation map (showing co-authoring) can spotlight authors who don't, but perhaps should, team up. Factor maps are used to reduce lengthy lists and show how selected items from one list cluster with each other based on principal components analysis. For instance, the top couple hundred keywords for a dataset can be reduced to 15 or so principal components to perceive main topical emphases and their interrelationships. In PCA, maps are based on multidimensional scaling and the axes do not matter. The links shown matter most for the relationship.

Design and Analysis Phase

Basic Analyses

The basic analytic tools in *VantagePoint,* lists and matrices, enable the analyst to become familiar with the data and identify priority areas for attention, as well as sort into meaningful combinations. These include simple activity counts of publications, patents, citations or keyword terms in different fields for a particular discipline, sub-discipline, country or over specific time periods. Higher activity counts serve as initial indicators and signal areas where further attention can be paid in order to identify emerging technologies. Similarly, the comparison of term frequencies in a particular subset with the frequency in the group as a whole points to areas of heightened levels of activity. Typically text mining information is highly skewed, with a small number of records providing the most pertinent information. This enables the analyst to choose cut-off points to get sub-sets data for subsequent analyses. Alternatively other mechanisms may be used for selecting thresholds, e.g. levels used in the published literature, domain expertise or intuition.

The data can be used directly or disaggregated into time slices to examine trends. Time slices enable one to understand trends by plotting changes in the number of publications, institutions, journals, conferences, countries or descriptors over different time periods. Indicators that change over time are selected for further analyses. Profiles get selected information out of multiple fields, for example, authors, keywords and publication year help in the identification of dominant researchers, research institutions and the institutional level at which research is taking place, whether in academia, industry or government laboratories.

Co-occurrence matrices are used to identify the frequency of shared terms, and can yield information on, for example, which researchers or institutions are working with each other, and the common issues. Network analyses identify nodes of concentrations of topics, authors, institutions etc. and the linkages with other groups. The existence of few hubs with enhanced communication and collaboration among participants because of the wealth of contacts and experience are signals of areas of heightened activity. Dimensional analyses using PCA to consolidate subtopics based on keyword co-occurrence patterns reduce large volumes of data to more manageable groupings that reflect major activities. PCA or PCD identifies groups of keywords that account for most of the variance in the dataset. Clustering techniques group data based on a chosen measure of similarity; tree based techniques successively divide data into classes. PCA can be repeated on the top tier principal components (factors) as well as what is left over after the principal components are taken out. It is possible to create numerous two or three-way (and higher) relationships of terms that have the potential to answer different questions. The analyses can be performed on the terms in the selected fields or on the citations in the documents, for example co-citation analysis looks at the number of times different publications are referenced in the same document. Table 10 illustrates possible relationships that can be developed.

Table 10. Term by Term Relationships (Selective)

	Authors	**Citations by our documents**	**Citations to our documents**	**Topics (e.g. keywords)**	**Institutions**	**Location**
Authors	Teaming	Research community	Esteem	Expertise	Expertise	Expertise
Citations by our documents		Co-citation analysis			Self-reliance	Self-reliance
Citations to our documents			Knowledge transfer		Esteem	Esteem
Topics (e.g. keywords)				Cluster analysis (research thrusts)	Expertise	Expertise
Source		Core	Impact	Core		
Year	Currency	Research space	Research space	Currency	Engagement	Engagement

In general, analyses are based on the combination of different *VantagePoint* analytic tools, with several passes over the data to identify cut-off points, groups or clusters -- then drilling down to find other relationships. For example the unique terms used by an organization to describe its research activity and the frequency of reoccurrence could be identified. It may be possible to get an indication of emerging areas based on the times that terms first appear. Alternatively innovation indicators (analytic outputs based on the questions asked) can be developed or the results fed into other analytic strategies such as competitive technology intelligence (who is doing what), roadmapping etc .

Advanced Analyses

More sophisticated analyses delve further into the relationships or patterns among the data, address unobserved variables (constructs) and model data behavior. Constructs, which are related to observable measures help in understanding S & T behavior (research activity, motives) and in interpreting the observed data. Porter and Cunningham identify six key constructs: prestige (esteem), life cycle of ideas (issue attention); invisible colleges (schools of thought); learning (knowledge increments); knowledge structure (approaches, concepts ideas associated with particular disciplines) and knowledge production (continuity of organizational research emphases). Statistical techniques are used in building the models and either deductive or inductive approaches may be used to model data relationships. In deductive approaches the data are fit to preset models and the goodness of fit is evaluated. Inductive approaches develop models from the patterns that emerge from the data. The models may be either qualitative or quantitative; deterministic or stochastic; and static or dynamic.

Choice Phase

After data analyses, experts can be used to review the results and reports prepared at different levels of detail -- for example a paragraph, 1 or 5-page summary, or a more complete and extensive document. These include maps, graphs, tables or combinations of these. Maps show topical convergence, topic-to-application links, and technology platform emergence. With landscape or document maps of topics or terms, elevated areas visually depict high concentrations of terms while the smaller peaks represent less prevalent terms. In white space mapping, we can visualize mountains of high interest, with indications of less intense activity.

*CHAPTER FIVE : EXAMPLES OF EMERGING TECHNOLOGY APPLICATIONS USING **VANTAGEPOINT***

Most applications of *VantagePoint* take place within organizations. Many of the several dozen studies that TPAC has done for other organizations are therefore not available for open discussion. Some of TPAC's emerging technology studies include:

a. Analysis of Emerging Technologies class projects
b. Bio Opportunities for Georgia Tech
c. Biochips for the Army
d. Synthetic Lumber alternatives
e. Emerging Technologies for Georgia Tech High Tech Indicators
f. Ceramics for Automotive Engine Application
g. GT Nano Analyses

This chapter presents a synopsis of these works and shows in detail the approach used in the last three works listed above as illustrative examples of the use of VP to support analyzes of national competitiveness, sectoral maturity, and institutional comparative position.

Analysis of Emerging Technologies Class Projects

From 1991 through 2001, Alan Porter annually taught "Analysis of Emerging Technologies" ("AOET") as graduate classes at Georgia Tech. Classes were taught for Industrial Engineering & Public Policy grad students, and for Management of Technology Executive Masters students. He also did a parallel AOET graduate course for the National Technological University (NTU) each year. In 2002 he taught AOET at the Technical University of Delft.

Students, either individually or as pairs, did a term project that consisted of a technology forecast and assessment of an emerging technology. On the order of 400 were thus generated. These varied widely in quality. Most of the NTU projects also served as company reports, as nearly all of these M.S. in MOT students worked full time for large companies (that financed their participation). These and a portion of the GT projects exemplify high quality emerging technology analyses. Most entailed application of *VantagePoint* (or its predecessor Technology Opportunities Analysis) software. A select sample remains available in electronic form. Illustrative titles include:

- Electronic cash
- Nickel-Metal Hydride Batteries for Electric Vehicles
- Video Image Processing for Intelligent Transportation Systems
- Internet-based Virtual Medical Records
- Protein Engineering and Heart Disease
- Speech Recognition in Call Centers

Bio Opportunities for Georgia Tech

TPAC's early work (1990-91) for Georgia Tech supported the Vice-President for Strategic Planning. He named these analyses, "Technology Opportunities Analysis," to convey the aim of identifying R&D opportunities. With hindsight, our biggest success was in identifying the potential for GT to expand its "bio" research and teaching. We identified relative national funding trends and counter-posed these against GT award trends. We also examined research projects and personnel capabilities relative to other research areas. TPAC surveyed GT researchers on their assessment of opportunity areas. We determined that bio-medical-engineering intersections had especially high potential for Georgia Tech.

Finding of the analyses were reported to the President of the university and its research leaders. A core of about 8 GT bio researchers used our findings to help justify increased commitment and investment to bio-medical-engineering. Within a few years, a major Whittaker Foundation grant spurred recruiting. Research flourished; 3 magnificent buildings have risen, and a joint PhD program in this area with Emory University is thriving. A leader of the bio group is now Dean of Engineering.

Biochips for the Army

TPAC performed a set of three analyses for the U.S. Army Medical Research Organization (Edgewood Arsenal, MD). The most interesting concerned the emerging technology of integrated, multi-function chips. These combine gene chips with MEMS with micro-scale chemistry. They build upon base silicon processing and chip production processes to enable high cost-effectiveness. We identified special potential for sensors and on-site analytics. We analyzed MEDLINE, INSPEC, and patent searches to assess the emerging trends and potentials.

Synthetic Lumber

An enterprise was considering investment in a new synthetic wood technology. This would entail very large capital investment in a manufacturing plant. We searched for alternative technologies in EI Compendex (also called ENGI through some providers), INSPEC, IPST (PaperChem), Science Citation Index (now included in Web of Knowledge), Business Index, and U.S. Patents. We were able to compare alternative technologies' maturation pathways and prospects. These helped the enterprise assess the risk in investing in the given synthetic lumber technology.

Emerging Technologies for Georgia Tech High Tech Indicators-HTI

TPAC explored measures of national R&D activity in emerging technologies (A.L. Porter et al., 2002). Our intent in this work was to improve the predictive validity of the "Technological Infrastructure Indicator', one of the input indicators we use to compute Georgia Tech's HTI that appears in the NSF's Engineering Indicators of technological competitiveness (http://www.nsf.gov/statistics/seind04/). We began with the latest American "critical technologies" analyses conducted by the RAND Corporation. We based our operationalization of emerging technologies on their set of "over the horizon" technologies. We used INSPEC and EI Compendex class codes to examine 33 countries' publications. Taking the class codes that correspond to the RAND technologies, we screened these to tally those showing strong recent, and increasing, R&D publication rates (which we called 'hot' areas). We operationalized this by computing two metrics. Firs, we included only technology class codes for which some 10%, or more, of the total articles occur in the most recent full year (or those published in journals or presented at conferences since 1969 for INSPEC and since 1970 for EI Compendex). Second, we calculate the ratio of publications in a technology category in the most recent full year (1999 at that time) to those three years earlier. To do this, after scanning our emerging technology categories, we found a ratio of at least two to be an effective screen.

Our measures resulted in coverage of the following five emerging technologies (we did not include energy technologies). In parentheses we indicate an example class code chosen as recently highly active and increasing:

- Software
- Computer hardware (semiconductor devices)
- Communication technologies (optical communications)
- Advanced materials for computing/communication technologies (rare metals – sum of silicon, tellurium & zirconium)
- Biotech (biological materials)

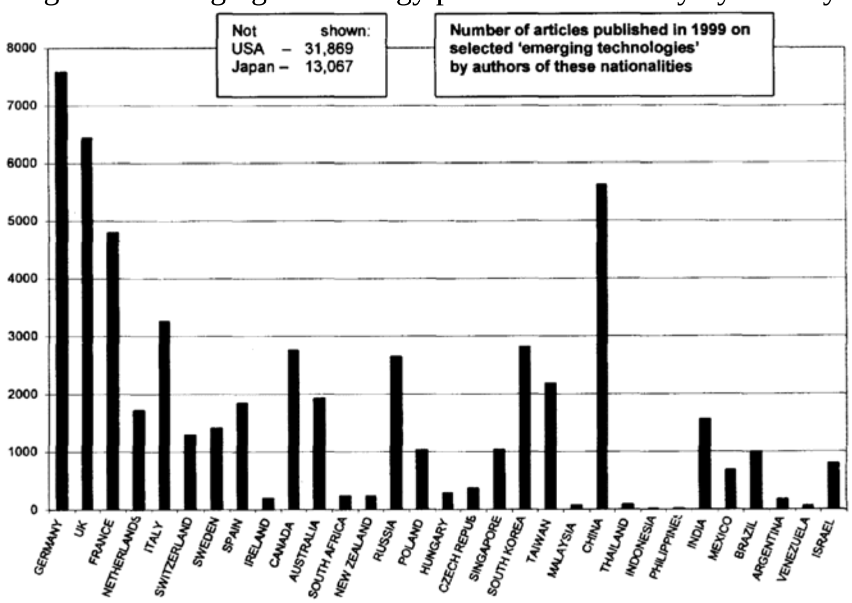
Based on expert judgment, we excluded some obvious class codes of certain EI Compedex that differed in nature with the technology categories where they appeared even though they met the dual criteria described. Similarly, we included Software, even though the growth criterion was violated, since this sector seemed a vital emerging technology domain at the time.

After cleaning the data, particularly relating to the country field for which we developed a set of thesauri, we computed the metrics and were able to identify a set of characteristics based on country specialization, relative weight and dynamism for each emerging technology and for all five emerging technologies identified. The resulting metrics showed strong convergence. Countries tended to be active, or not, in most of the categories generally. The measures pointed to China as a rising emerging technologies R&D power. Certain other nations evidenced a striking lack of R&D activity in these areas, posing questions about their longer range high-tech competitiveness.

Figure 6 depicts the results for the sum of the five emerging technologies leaving out the "research superpowers," the U.S. and Japan, to improve scaling. The most striking observation concerned China's strong presence. This figure suggested we could identify tiers in research activity on emerging technologies:

- Superpowers: USA and Japan
- Research Powerhouses: Germany, UK, China, and France
- Strong Players (those with over 2,000 annual publications): Italy, South Korea, Canada, Russia, Taiwan
- Solid Presence: 11 countries with 670-2000 annual publications –Australia, Spain, the Netherlands, India, Sweden, Switzerland, Singapore, Poland, Brazil, Israel, Mexico
- Laggards (those with about 200-400 annual publications): the Czech Republic, Hungary, New Zealand, South Africa, Ireland, Argentina
- Those lacking critical research mass (<100 publications): Thailand, Malaysia, Venezuela, Indonesia, and the Philippines.

Figure 6: Emerging Technology publication Activity by country



Consistent with the HTI, we preferred to present total national activity as opposed to weighted data (normalized by number of local scientists or engineers, or per capita) since the former reflects better national capabilities for export competitiveness.

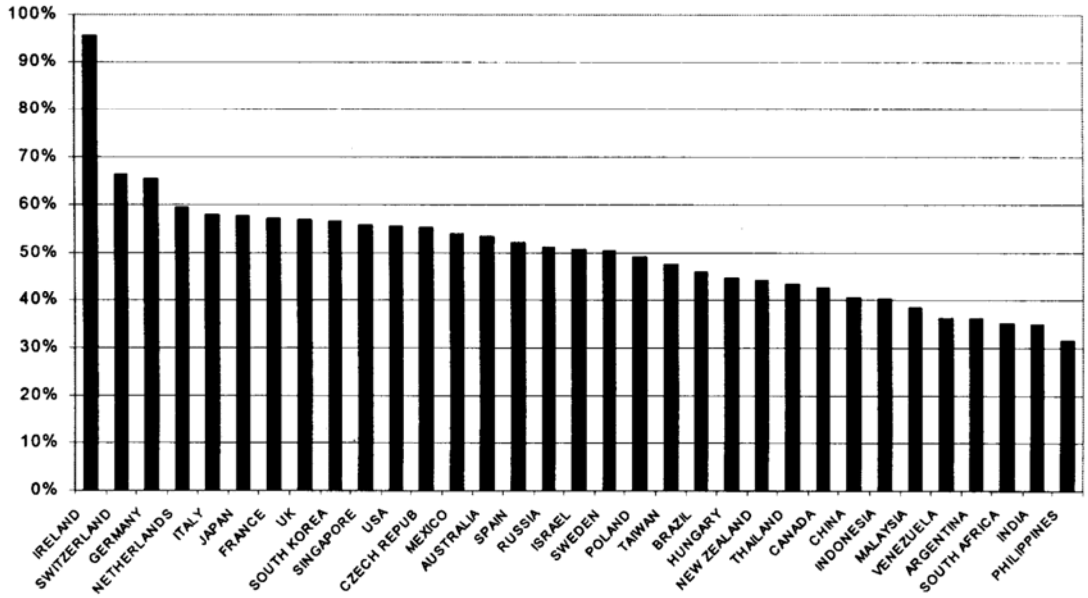
In addition, to make it easier to compare countries' relative performance, we used the 'S-scores' scaling approach, which consisted on scaling the 33 countries on a relative basis with the leader on a particular variable as '100' and the last country as '0', hence identifying for each country their fraction of the highest country value (e.g. in average, for the five technologies, Japan represents 46% of the USA, the leading country in all technologies and so on). Table 11 presents both ranking for each technology and average S-score for 1999.

Table 11: National 'Emerging Technology' R&D publication ranks
and average S-scores for 1999

Ranks	ENGI	Optical comm (ENGI)	Comp hdwr (ENGI)	Semi mtls (ENGI)	Biotech (ENGI)	Software (INSPEC)	Average S-score (5 ETs)
USA	1	1	1	1	1	1	100.00
Japan	2	2	2	2	2	2	45.99
Germany	4	3	3	3	4	4	24.73
UK	5	5	4	6	3	3	19.49
France	6	6	5	5	6	6	15.78
Netherlands	15	15	15	16	10	13	5.32
Italy	8	8	9	10	9	7	10.28
Switzerland	19	17	13	17	17	15	4.08
Sweden	16	18	12	15	11	14	4.82
Spain	14	11	15	13	15	11	6.05
Ireland	28	28	25	27	28	23	0.60
Canada	7	10	10	11	5	8	8.51
Australia	13	14	18	19	12	9	5.52
South Africa	24	27	24	26	26	27	0.83
New Zealand	26	26	29	28	24	22	0.73
Russia	9	7	11	7	8	18	8.96
Poland	18	16	19	14	20	21	3.51
Hungary	25	24	25	24	23	26	1.02
Czech Repub	23	23	23	23	25	25	1.19
Singapore	20	20	17	20	22	17	3.38
South Korea	10	9	7	9	16	10	9.40
Taiwan	11	12	8	12	14	12	7.20
Malaysia	30	29	27	30	31	30	0.21
China	3	4	6	4	7	5	17.48
Thailand	29	31	29	31	29	28	0.26
Indonesia	32	32	32	32	31	33	0.11
Philippines	33	33	33	33	33	32	0.03
India	12	13	14	8	13	20	5.99
Mexico	22	21	21	22	21	24	2.29
Brazil	17	19	20	18	19	16	3.25
Argentina	27	25	28	25	27	29	0.69
Venezuela	31	29	31	29	30	31	0.26
Israel	21	22	22	21	18	19	2.51

Finally, we also reported emerging technologies emphasis by country. To do that, we computed the percentage of a country's engineering research in the five emerging technologies identified. However, given the categorization issues involved, and the small number of publications for some countries, the analysis of such results required caution. Figure 7 reproduces such analysis.

Figure 7: Emerging Technology emphasis by country.



An emerging technologies indicator is now included in our statistics-only version of the Georgia Tech High Tech Indicators prepared for the National Science Foundation's *Science & Engineering Indicators.*

Ceramics for Automotive Engine Application

This study was led by Bob Watts of the U.S. Army with Alan Porter collaborating (Watts & Porter, 1997). The purpose of the study was to address a possible technological substitution for the U.S. Army –use of ceramics in place of steel in tank or automotive engine components. Analytical approaches, including the use of "Keyword Richness" to flag a significant step change in ceramics technology maturation, were reported.

The way the authors approached this exercise included the following steps:

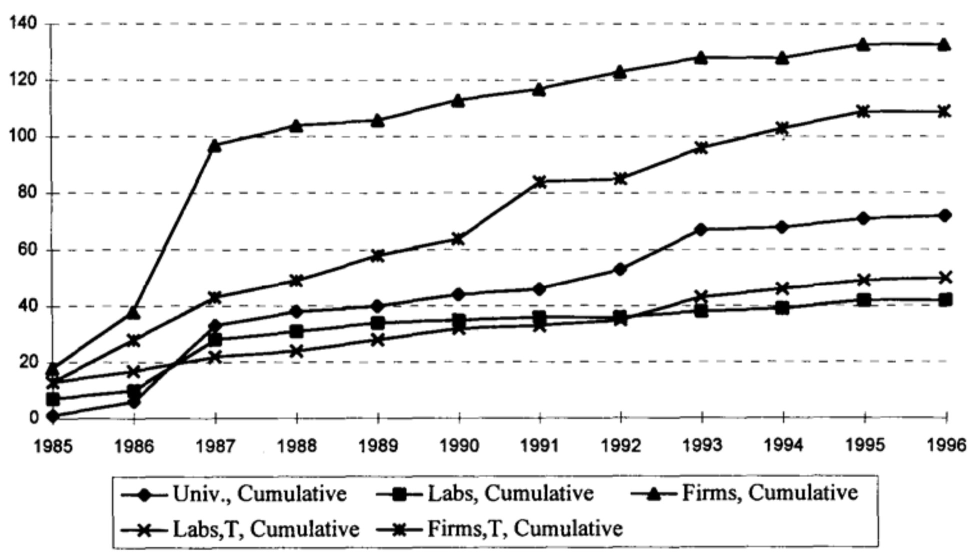
1. Search on the basic topical term(s) in multiple databases.
2. Download electronic abstracts from a prime, available database; examine cumulated keywords, etc., to refine topic understanding to generate a good search algorithm.
3. Redo search in most advantageous database(s); download abstracts.
4. Examine keywords, title words, and abstract words and phrases; read abstracts to gain fluency with related activities, applications, key players, dispersion.
5. Plot trends in overall activity, topic-specific activity, institution-specific activity, etc.
6. Consider activity patterns by type (academic, government, industry) or other delimiters of interest.
7. Model the technology life cycle.
8. Cluster technological or other activity associated with the target.

9. Map key supporting technologies, institutional interests, etc.
10. Depict maps at different time slices.
11. Map likely future technological or competitive profiles, if appropriate.
12. Develop a technology decomposition tree, including tagging players; breakout for key contributing technologies.
13. Perform analyses on special areas (e.g., gap analysis).

A preliminary search (Step 1) located prior forecasts, in particular, a Delphi study. The Delphi respondents had identified enabling technologies and application barriers that existed in the mid-1980s. These provided good leads for further bibliometric searches on both the enabling and primary technologies from *Engineering Index* and *U.S. Patents* (Step 3). The main search addressed 1985 to 1995 on "ceramic" within 6 words of "engine." The resulting search records were downloaded in electronic form and subdivided into two files--turbine and other. Turbines (file of 214 records) provide a possible lead technology indicator. For some purposes, the files were further pruned to include only records from the top 100 institutions--universities, government labs, and commercial films--publishing on ceramic engines.

Figure 8 shows the chronology of the publications for the three source groupings for the two categories of ceramic engine publications at the time of the analysis (There were only three turbine abstracts from universities).

Figure 8: Ceramic engine publication



Abbreviation: T=Turbine

Table 12 was created to show the co-occurrence matrix for the government laboratory organizations that produced the most non-turbine publications and the number of matches of the most frequently used keywords. Similar tables were compiled for academia and industry. These tables identified those who were most active in publication of ceramic engine R&D.

Table 12: National Ceramic Engine Keywords

	Inst natl des hydrocarbures et de la chimie	Oak Ridge Natl Lab	NASA	Acad of Sciences	Battelle Memorial Inst	Government Industrial Research Inst at Nagoya	Chinese Acad of Sciences	Lawrence Berkeley Lab	Commission of the European Communities	Argonne Natl Lab	Southwest Research Inst	Admiralty Research Establishment	Acad of Sciences of the Ukrainian SSR	Battelle	Government Industrial Research Inst	Battelle Columbus Lab	Federal Dep of Science & Technology	Total
Ceramic materials	6	6	4	5	3	4	3	0	1	2	4	3	0	4	1	1	3	50
Diesel engines	3	6	2	3	2	1	0	7	1	1	1	0	2	0	3	0	0	32
Tribology	2	1	0	1	2	3	1	0	2	1	1	1	0	0	0	0	0	15
Mechanical properties	2	1	2	0	1	0	2	0	0	0	0	0	1	0	0	0	0	9
Heat engines	2	2	2	0	0	0	0	0	0	1	0	1	0	0	0	0	0	8
Thermomechanical ceramic	2	0	1	0	1	0	1	0	0	0	0	0	0	0	0	2	0	7
Physical properties	2	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	6
Ceramic products	2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Silicon nitride	2	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	6
Superchargers and supercharging	0	0	0	1	0	1	1	0	0	2	0	0	0	0	0	0	0	5
Wear of materials	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	5
Materials testing	2	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	5
Impact damage	0	1	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	5
Impact testing	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	4
Grinding (machining)	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Microcrack	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	4
Durability	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	4
Cracks	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	3
Protective coatings	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	3
Scanning electron microscopy	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	3
Air engines	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3
Surface properties	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	3
Degradation	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2
Rocket engines	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
Total	31	28	18	15	14	14	14	12	9	8	7	5	4	4	4	3	3	

The co-occurring keywords helped the identification of the areas of concentration. It revealed the balance of development cycle participants, with industry taking a strong lead in applied research and development. For less mature technologies (i.e., electrorheology or artificial intelligence), a greater proportion of activity by basic research institutions and lower activity from influential sponsors was observed. The abstracts also revealed a balance of R&D activity across the industry infrastructure (i.e., components, engine, and vehicle manufacturers).

The literature analyzed conveyed that the advantages of ceramic components had begun to be proven; the technology was maturing. The three most-cited barriers included cost, material properties verification, and coating and bonding technologies--three candidates to explain the 1987 surge in publication activity shown in Figure 8 above. Using the terms ceramic adjacent to coating or ceramic adjacent to bonding yielded 234 related patents during the 1980 to 1995 period. Figure 9 depicts the chronology of patents issued and the cumulative patent growth in the ceramic coating and bonding field. The significant rise in number of patents issued in 1986 and 1987 provided an explanation for the industrial publication surge in 1987. The cumulative ceramic coating and bonding patents were modeled (Step 7) by three Fisher-Pry equations, each with a different technology growth limit (i.e., 350, 450, and 550 patents). The growth limits were selected because limits below 350 patents and above 600 patents provided lower coefficients of determination. These equations were then used to generate patent forecasts through the year 2005, as shown in Figure 10.

Figure 9: Cumulative ceramic coating and bounding patents.

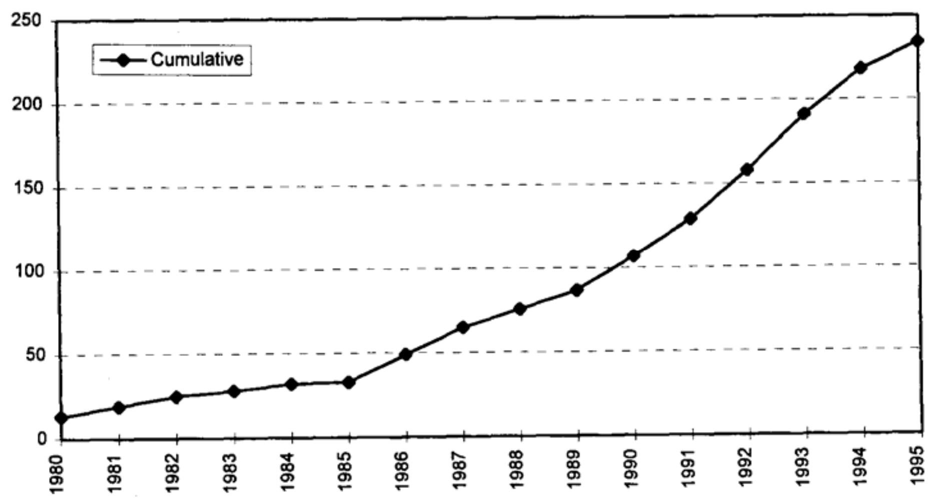
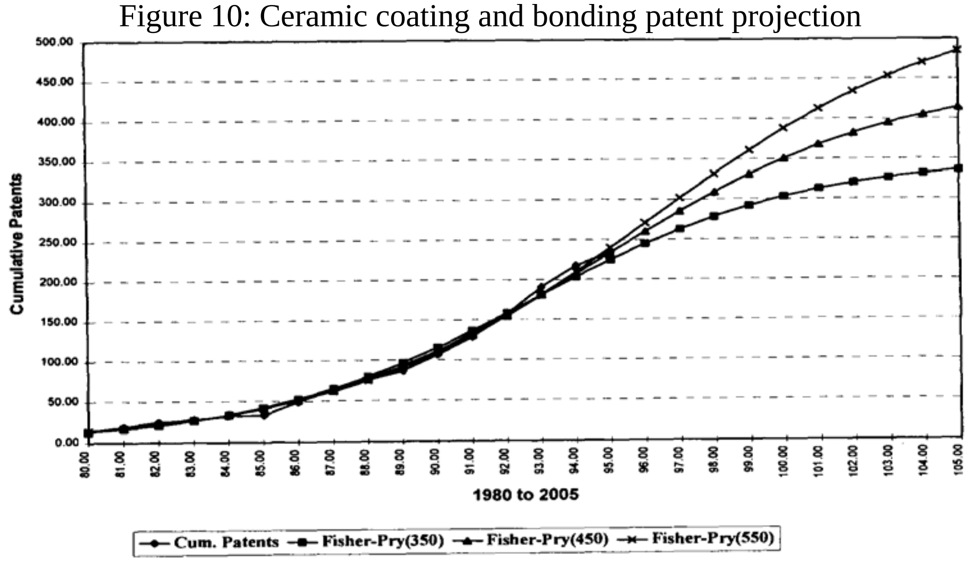


Figure 10: Ceramic coating and bonding patent projection



To extend the maturity analysis from enabling technologies (i.e., ceramic coating and bonding) to ceramic engine technology more generally, two bibliometric approaches were applied. The 100 most-used keywords from the 426 nonturbine ceramic engine abstracts were subdivided into two groups: material types and a combined group of material properties and applications. We then generated a co-occurrence matrix--materials versus properties and applications (Table 13). Two observations from this table, in regard to ceramic engine technology, included the apparent emergence of silicon nitride as the ceramic material of choice and the presence of competing materials (e.g., aluminum compounds, metal matrix composites, metals and alloys, superalloys).

Table 13: Ceramic Engine Materials versus Material Properties/Applications

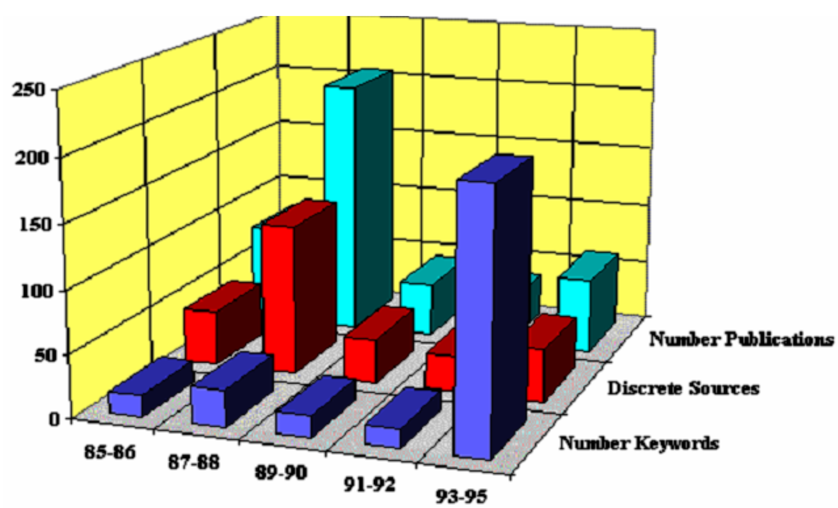
	Silicon nitride	Composite materials	Aluminum compounds	Metallic matrix composites	Metals and alloys	Superalloys	Ceramic heat engines	Aluminum titanate	Ceramic fibers	Ceramic matrix composites	Low-cost ceramic composite	Metallic silicon alloys	Ceramic coatings	Titanium alloys	Titanium oxides	Yttrium	Zirconia
Properties	14	2	1	1	1	1	0	1	0	0	0	1	0	1	0	0	0
Strength	15	0	1	0	0	0	1	1	0	1	1	0	0	0	0	1	0
Applications	5	2	0	1	0	2	0	0	1	0	0	1	0	0	0	0	1
Application	5	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0
Performance	2	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0
Process	4	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0
Structural	4	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Piston	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Parts	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Coatings	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Processing	4	1	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
Production	6	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
Valve	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Reliability	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
Coating	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
Characteristics	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Problems	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Automotive	2	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Cost	1	2	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0
Economy	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Quality	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Processes	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Durability	3	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0
Fatigue	2	1	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0
Manufacturing	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
Liner	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pistons	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Turbocharger	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Valves	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

To obtain a temporal perspective on the types and usage of keywords related to ceramic engine technology, the nonturbine ceramic engine abstract file was subcategorized into five 2-year periods of publication abstracts. Co-occurrence matrices of sources versus keywords were generated. Table 14 summarizes the co-occurrence matrices by defining the level of activity (e.g., the number of discrete publication sources and associated number of publications) and the level of focus of the documented research (e.g., the number of discretely different keywords). The evolution of a technology was shown by creating Figure 11, which depicted Table 14 data, and was considered in terms of the Utterback and Abernathy model on product and process innovations, that is, the assumption that early research is product focused and attracts many industry participants. According to this model, once a dominant design emerges, research shifts towards process technology, and the number of industry participants declines. As shown in Table 14, in the 1987-1988 period, the level of interest in the technology peaked as indicated by the numbers of publications (207) and participating organizations (120). The areas of R&D, however, were quite focused, as indicated by the number of different keywords used (29). Contrast this profile with that for 1993-1995: far fewer participating organizations (42), a proportional reduction in number of publications, but tremendous expansion of the detail and issues addressed (201 different keywords used).

Table 14: Co-ocurrence Matrices Summary

Years	Discrete sources	Number of publications	Number of keywords
1985–1986	44	79	17
1987–1988	120	207	29
1989–1990	35	44	17
1991–1992	29	36	15
1993–1996	42	60	201

Figure 11: Technology maturity and keyword diffusion



To see the evolution of the types of technological activities addressed over the time periods, the common keywords across periods were eliminated. Table 15 was created to present the chronology of the use of the remaining words. Innovation sequences often start with an invention (e.g., technology application such as the invention of the internal combustion engine), followed by the emergence of related sciences (e.g., tribology, combustion, etc.). Based on Table 15, which shows that the ceramic engine technology terms have evolved toward analytic sciences in addition to expanding to processes, material properties verification, and application fields, we were able to support the notion of a maturing technology poised to assume niche positions in specialty material growth markets.

Table 15: Technology Maturity versus Keyword Usage

Report period	Generic material	Generic application	Needs/function	Specific material	Specific application
1985–1986	Materials, metallic silicon alloys, refractory materials	Air engines, sensors, superchargers and supercharging	Lubricants		
1987–1988	Materials, metals and alloys, refractory materials, superalloys, composite materials	Machine components, superchargers and supercharging	Protective coatings, waste heat utilization	Silicon nitride, silicon carbide	Combustors, domes and shells
1989–1990	Composite materials	Gas engines	Friction materials, heat transfer	Silicon nitride	Domes and shells, valves and valve gear
1991–1992	Metals and alloys, nonmetallic materials			Zirconia	Pistons
1993–1995	Porous materials, aluminum compounds, amorphous materials, metallic matrix composites, composite materials		Thermal insulating materials, lubricants, thermomechanical ceramic	Silicon nitride, ceramic fibers, aluminum titanate	Seals, ceramic heat-insulated engine, braided ceramic fiber seals, adiabatic engines
	Generic approach	**Specific approach/process**	**Material characteristic verification**	**Enabled technology**	**Analytical science**
1985–1986					
1987–1988	Powders, automotive engineering, powder metallurgy				
1989–1990		Castings			
1991–1992			Microscopic examination	Hydrogen fuels, methanol, diesel fuels (alternative fuels)	
1993–1995		Sintering, braiding process	Microstructure, strength of materials, material testing, physical properties, volume fraction, high temperature properties, mechanical properties, fatigue testing, reliability, wear of materials, durability, axial/circumferential strength, creep, defects	Aromatic polyphenyl ether type oil	Mathematical models, tribology, finite element method, computer simulation, computational geometry

Ceramics Automotive Applications

Ceramic components	Material	Supplier	User
Intake and exhaust valves	Silicon nitride	Ceramtec Div. of Hoechst A.G.	Daimler Benz
Exhaust portliner	Aluminum titanate	Ceramtec Div. of Hoechst A.G.	Porsche A.G.
Brake engine retarder master piston wear pad	Silicon nitride	ENCERATEC, Inc.	Cummins N14 Engine
Cam roller follower	Silicon nitride	Kyocera Corp.	Detroit Diesel (Series 50)
Ceramic tappet	Sintered silicon nitride	NKG Spark Plug Co.	Nissan Diesel Motor Co.
Ceramic coatings		Ceramics Corp. of America (Cercoa)	
Ceramic coatings	Zirconia coating with strain isolator	Technetic's Corp.	

Given the lack of publicly available information, we were unable to study in detail the other two application barrier issues (cost and manufactured material property verification). Component cost data were sought through both literature review and phone contacts with material journal publishers and ceramic engine component manufacturers. These efforts uncovered the fact that ceramic component cost data represented confidential information between component suppliers and end-item manufacturers (e.g., automotive and engine). A search of *U.S. Patents* using the terms <u>ceramic material quality</u>, <u>ceramic non-destructive test</u>, and <u>ceramic property test</u>, uncovered only four relevant patents. Thus, the fruitless component cost and patent searches, along with commercialization announcements, supported one conclusion: the manufacturing costs and process verification techniques were being held secret to obtain and maintain competitive advantage.

A detailed analysis of the implications of the findings reported in the framework of this study is publicly accessible in Watts &Porter 1997.

To update the story, after being convinced of the heightened potential, the Army Tank-Automotive Research, Development & Engineering Center (TARDEC) sought expertise in thin-film ceramics. "Tech Mining" played an essential role in identifying that such research was taking place in a quite different ceramics venue – namely, semiconductor (integrated circuit) R&D. This led to identification of two key R&D groups and Army funding of $million projects with Sandia National Lab and a private company. The focus was to adapt vapor deposition of ceramics to coat automotive engine components. In 2004, the Rouge River plant began operations to coat Abrams tank turbine blades with ceramics.

<u>Recent GT Nano Analyses</u>

TPAC has been profiling Georgia Tech research in nanotechnology -- broadly considered an emerging technology - using text mining techniques, as a starting point for benchmarking the institution's research against that of other organizations. For the analysis we used records abstracted from R&D technical journal articles from EI Compendex and INSPEC databases for the period 1995 to 2005 (now incomplete).

We used a broad, inclusive search to find Georgia Tech "nano" publication records in the two databases. The exact text of the search, which is directly usable from the EI Village "Expert Search" screen is as follows:

```
((nano* WN ALL) AND
(((georgia OR ga) AND (tech OR technol*)) WN AF) AND
(1995-2005 WN YR))
```

The "nano*" term will match all text strings beginning with "nano". The "WN ALL" is a search tag that will match the text in all fields in the database (title, abstract, keyword, etc.) The second grouping searches for variants of "Georgia Tech" within the "Affiliation" field of the database (using the "WN AF" tag). The term <u>tech</u> is not truncated with a wildcard so as not to match variants of <u>technical</u> which also may appear in affiliation names. We used a wildcard in the <u>technol</u> term to be sure we match the abbreviation <u>technol</u> as well as <u>technology</u>. The last string in the search searches for records of articles with publication dates in 1995 to 2005. The databases offer an “auto stemming” feature which will automatically search for other forms of search terms such as plurals. This feature was left off in our search because the “nano*” search covers all plural forms of matched terms. Furthermore, leaving this feature enabled in our search matched the “ga” term to “gas” and was adding records published from non-GT organizations as a result.

The search was performed in EI Compendex and INSPEC on September 14, 2005. Table 16 shows each database and the number of publication records retrieved.

Table 16: Search results as of 9/14/05

INSPEC	587
EI Compendex	688
Combined (OR)	1275
Unique Total	**895**

The 587 INSPEC records and the 688 EI Compendex records were downloaded and imported separately into VantagePoint. We keep the data from the two databases separate until after importing because there are slight variations in the record formats which would cause INSPEC records not to import correctly through an EI Compendex filter, and vice versa. Once the records are imported to VantagePoint we save the INSPEC and Compendex as separate VP files.

The VantagePoint files just created can be fused quite easily into one VP file using the Dataset Fusion tool. The VP files are more easily fused than the raw “tagged record” files because the Import Filters are not used. Fields are combined according to field name, and the VP user has control over which fields from each dataset are combined. The new dataset contains 1275 records from both databases.

There is considerable overlap in Compendex and INSPEC journal coverage which means many of these 1275 records are likely to be duplicates. For this analysis we consider two records duplicates if their titles match exactly. On this sole criterion, 280 duplicate records are identified and removed from the dataset. **The total number of unique records from INSPEC and Compendex is 895**. These data will be used for the analysis of Georgia Tech “nano” research.

After the records have been imported to VantagePoint, we can begin manipulating the data to profile Georgia Tech research groups and their publication interests to date. Given the broad spectrum of nanotechnology we will focus on comparing the work of only one research group at Georgia Tech. But how do we decide who to profile?

Again we turn to VantagePoint. We clean the “Authors” list using the List Cleanup tool and create the new Field “Authors-Cleaned”. We create a group of the 50 authors for whom we have captured 10 or more records. The list of these top 50 Authors appears in Table 17. It should be noted that appearance in this list does not guarantee that the researcher is from Georgia Tech. A non-GT researcher's name may appear on this list if they have co-authored 10 or more papers with a GT faculty member.

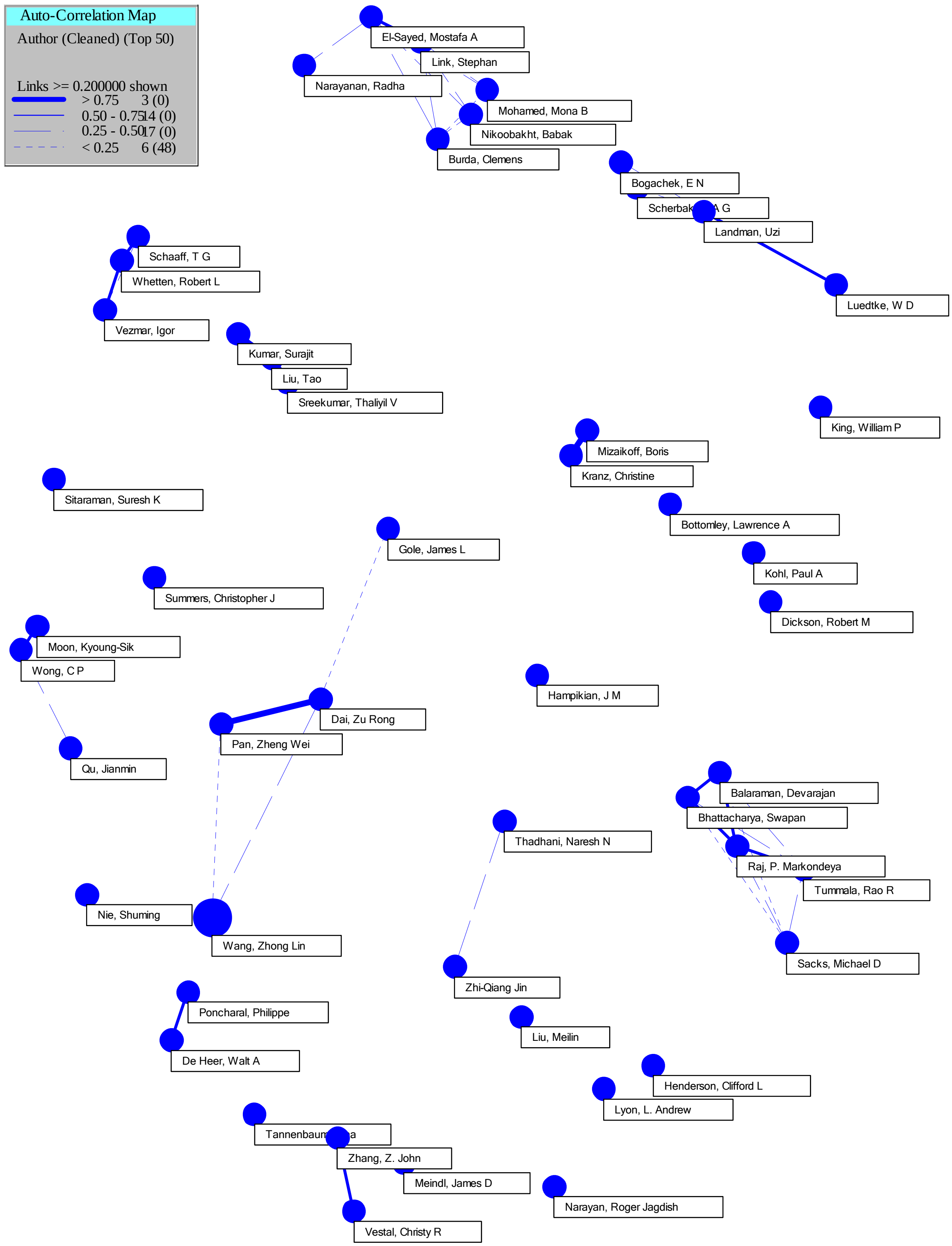
Table 17 Top 50 Georgia Tech "Nano" Authors

# Records	Author	# Records	Author
146	Wang, Zhong Lin	13	Lyon, L. Andrew
20	Gole, James L	13	Mizaikoff, Boris
17	Dai, Zu Rong	13	Nie, Shuming
10	Pan, Zheng Wei	13	Summers, Christopher J
71	El-Sayed, Mostafa A	12	Balaraman, Devarajan
45	Landman, Uzi	12	Liu, Tao
40	Wong, C P	12	Poncharal, Philippe
33	Tummala, Rao R	11	Bhattacharya, Swapan
30	Link, Stephan	11	Bogachek, E N
28	Zhang, Z. John	11	Hampikian, J M
27	Whetten, Robert L	11	King, William P
25	Burda, Clemens	11	Kohl, Paul A
25	Kumar, Surajit	11	Mohamed, Mona B
24	Raj, P. Markondeya	11	Moon, Kyoung-Sik
20	Sitaraman, Suresh K	11	Narayanan, Radha
20	Thadhani, Naresh N	11	Nikoobakht, Babak
19	De Heer, Walt A	11	Scherbakov, A G
19	Dickson, Robert M	11	Vezmar, Igor
19	Liu, Meilin	10	Bottomley, Lawrence A
16	Luedtke, W D	10	Meindl, James D
16	Narayan, Roger Jagdish	10	Qu, Jianmin
16	Sacks, Michael D	10	Schaaff, T G
15	Tannenbaum, Rina	10	Sreekumar, Thaliyil V
13	Henderson, Clifford L	10	Vestal, Christy R
13	Kranz, Christine	10	Zhi-Qiang Jin

One way that we can represent research groups is by using a correlation measure to track co-authorship frequencies. The Auto-Correlation map does this using a Pearson's R correlation function. Author names are represented as map nodes and links between the nodes indicate co-authorship. The line pattern gives a relative measure of the frequency of co-authorship More information about an author appearing in the map is available in dropdown boxes. We can quickly view an author's most frequently occurring keywords or journal of publication. . We will examine in detail research group led by Dr. Z.L. Wang.

It is undesirable to show all co-authorship links in a map such as this because so many lines would make the map unreadable. An author as prolific as Dr. Wang has co-authored papers with many more researchers than the map indicates. As a compromise we set the preferences in this map to show only links for correlations >.20.

Figure 12 Auto-correlation Map (Co-authorship) map showing Keywords in dropdown boxes.



Looking at Figure 12 we can see that there are 2 authors that appear closely connected in publication to Dr. Wang. We will add these two authors, (Dai and Pan) as well as James Gole, who is connected to Dai to what we will call the "Wang Group".

We created a sub-dataset of all publication from each of these 4 authors in the Wang Group. This new dataset contains 159 publication records. We can use these records to gain perspective on publication interests by looking at author keywords and their frequencies over time. We obtain the keyword frequencies for each year in VantagePoint by creating a Keyword X Year co-occurrence matrix.

For this analysis we have defined 3 categories of keywords.

- "Top" Keywords – Wang Group's most frequently occurring keywords from the 159 publications from all years (1995-2005)
- "Hot" Keywords – Keywords that appear 5 or more times from 1995 to 2005, with 75% or more of those occurrences in 2004 and 2005 (combined). High frequencies of a keyword in recent years may suggest a start-up effort in a new direction.
- "New" Keywords – Keywords appearing for the first time in the year 2005.

Due to publication lag as well as changes to database controlled vocabulary these keywords may not actually be "new" or "hot" but could provide clues to this group's most recent publication interest.

Table 18 – "Top" "Hot" and "New" Keywords for Wang Group

"Top" Keywords	"Hot" Keywords
Nanostructured materials	Optoelectronic devices
Transmission electron microscopy	Iron
Zinc compounds	permanent magnets
Nanotechnology	Piezoelectric materials
II-VI semiconductors	Praseodymium alloys
Scanning electron microscopy	boron alloys
Crystal structure	Sensors
Synthesis (chemical	Zinc oxide
Carbon nanotubes	Compaction
Electron diffraction	exchange interactions (electron)
Iron alloys	grain size
Photoluminescence	nanocomposites
Self assembly	Ferromagnetic materials
wide band gap semiconductors	Catalysts
	Single crystals

"New" Keywords	
Cooling	Partial pressure
Polycrystalline materials	Phase diagrams
Aluminum nitride	Phonons

Band structure	Polarization
Carbon monoxide	Polyethylenes
Electrophoresis	Polypyrroles
Fast Fourier transforms	Porous silicon
Feedback control	Problem solving
Filters (for fluids)	Random processes
Fracture toughness	Sensitivity analysis
Gallium nitride	Silicon wafers
Interconnection networks	Sol-gels
Layered manufacturing	Spectroscopic analysis
Magnetic fields	Tellurium
Metallorganic chemical vapor deposition	Thermoelectricity
Optical properties	Toughness
Organic solvents	Vapor deposition

We have created here a preliminary profile of a Georgia Tech research group. Before directly comparing research by this group to that of the whole of the research community it is prudent to confirm if this analysis is consistent with other sources such as the Georgia Tech Nanoscience and Nanotechnology Website. A robust, but targeted search is essential to a balanced assessment of this research group's standing among peers.

Discussion and Conclusions

We began this report with a review of the concept of emerging technologies and how it is defined in the discourse and literature on technological change. The review revealed that while the concept is widely used, it is seldom defined, and even less often measured.

Our review of methods explains why. The characteristics of rapid growth, newness, untapped market potential, and a high-technology base are quite difficult to put into operational form with the data sources available, especially all at the same time. Different data sources carry different elements of important information, and the search for emerging structures, even with the sophisticated information tools now available, must be supplemented with the search for additional information on a case-by-case basis in order to meet the goals of a monitoring system for emerging technologies.

Vantage Point's flexibility in analytic tools can be a powerful tool in such a monitoring system, especially at the stage of case-by-case analysis. The ability to search for structure within an area using several different data elements and several techniques is particularly useful, and the visualization features of the software make the results accessible to those not familiar with the system itself.

In conclusion, bibliometric techniques hold significant potential for use in monitoring systems, but also many limitations. At the current time, national monitoring for emerging technologies therefore must still depend on a combination of quantitative analysis with the knowledge of market-oriented technical experts.

REFERENCES

Abernathy, W. J., & Utterback, J. M. (1978). Patterns of industrial innovation. *Technology Review*(June-July), 40-47.

Allen, T. (1977). *Managing the flow of technology*: MIT Press.

Brooks, H. (1994). The Relationship between Science and Technology. *Research Policy, 23*(5), 477-486.

Bush, V. (1945). *Science: The Endless Frontier. A Report to the President of the United States Franklin D. Roosevelt*. Washington, D.C.: U.S. Government Printing Office.

Busquin, P. (2002). *The foresight dimension of the European research area.* Paper presented at the The Role of Foresight in the Selection of Research Priorities, Seville, Spain.

Carlsson, B., Jacobsson, S., Holmen, M., & Rickne, A. (2002). Innovation systems: analytical and methodological issues. *Research Policy, 31*(2), 233-245.

Carrocher, N., Malerba, F., & Montobbio, F. (2003). *The emergence of new technologies in the ICT field*. Retrieved October 12, 2005, from http://eco.uninsubria.it

Chapple, K., Markusen, A., Schrock, G., Yamamoto, D., & Yu, P. (2004). Gauging metropolitan "high-tech" and "I-tech" activity. *Economic Development Quarterly, 18*(1), 10-29.

Christensen, C. M. (1997). *The Innovator's Dilemma*: Harvard Business School Press.

Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution*: Harvard Business School Press.

Corrocher, N., Malerba, F., & Montobbio, F. (2003). The emergence of new technologies in the ICT field: main actors, geographical distribution, and knowledge sources. *Working Papers of Faculty of Economics Università degli Studi dell'Insubria, 37 Varese, Italy*, http://eco.uninsubria.it/dipeco/Quaderni/files/QF2003_2037.pdf.

Cuhls, K. (2003). From forecasting to foresight processes - New participative foresight activities in Germany. *Journal of Forecasting, 22*, 93-111.

Day, G. S., & Schoemaker, P. J. H. (2000). A different game. In G. S. Day & P. J. H. Schoemaker (Eds.), *Wharton on Managing Emerging Technologies.* New York, NY: John Wiley & Sons.

Edquist, C., & Hommen, L. (1999). Systems of innovation: theory and policy for the demand side. *Technology in Society, 21*(1), 63-79.

Eto, H. (2003). The suitability of technology forecasting/foresight methods for decision systems and strategy - A Japanese view. *Technological Forecasting and Social Change, 70*(3), 231-249.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. *Research Policy, 29*(2), 109-123.

FOREN. (2001). *A Practical Guide to Regional Foresight*: FOREN Network
European Commission Research Directorate General.

Foster, R., & Kaplan, S. (Eds.). (2001). *Creative Destruction: Why Companies That Are Built to Last Underperform the Market--And How to Successfully Transform Them*: Currency.

Freeman, C., & Soete, L. (1997). *The Economics of Industrial Innovation*. Cambridge Mass: The MIT Press.
Georghiou, L. (2002). *Commentary on Science and technology foresight in Europe: A prospective view*. Paper presented at the The Role of Foresight in the Selection of Research Priorities, Seville, Spain.
Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The dynamics of science and research in contemporary societies.*: Chapters 1-2. Sage Publications.
Halal, W. E., Kull, M. D., & Leffmann, A. (1998). The George Washington University Forecast of Emerging Technologies: A Continuous Assessment of the Technology Revolution. *Technological Forecasting and Social Change, 59*(1), 89-110.
Haugg, K. (2002). *Science and technology foresight in Europe: Could foresight be a misleading process*. Paper presented at the The Role of Foresight in the Selection of Research Priorities, Seville, Spain.
Hecker, D. (1999). High-technology employment: a broader view. *Monthly Labor Review, June 1999*.
Hung, S. C., & Chu, Y. Y. (in press). Stimulating new industries from emerging technologies: challenges for the public sector. *Technovation*, 1-7.
Islas, J. (1997). Getting round the lock-in in electricity generating systems: The example of the gas turbine. *Research Policy, 26*(1), 49-66.
Kassicieh, S. K., Kirchhoff, B. A., Walsh, S. T., & McWhorter, P. J. (2002). The role of small firms in the transfer of disruptive technologies. *Technovation, 22*(11), 667-674.
Keenan, M. (2003). Identifying emerging generic technologies at the national level: the UK experience. *Journal of Forecasting, 22*(2-3), 129-160.
Kline, S., & Rosenberg, N. (1986). An overview of innovation. In R. Landau & N. Rosenberg (Eds.), *The positive sum strategy: harnessing technology for economic growth*. Washington, D.C.: National Academy Press.
Kostoff, R. N. (2004). *Systematic acceleration of radical discovery and innovation in science and technology*, from http://www.onr.navy.mil/sci_tech/special/354/technowatch/textmine.asp
Kostoff, R. N., Boylan, R., & Simons, G. R. (2004). Disruptive technology roadmaps. *Technological Forecasting and Social Change, 71*(1-2), 141-159.
Kuhn, T. S. (1966). *The Structure of Scientific Revolutions* (Third ed.). Chicago: The University of Chicago Press.
Landry, R., Amara, N., & Lamari, M. (2002). Does social capital determine innovation? To what extent? *Technological Forecasting and Social Change, 69*(7), 681-701.
Langenhove, L. (2002). Science and technnology foresight in Europe: A prospective view. In IPTS-JRC (Ed.), *The Role of Foresight in the Selection of Research Priorities*. Seville, Spain: European Comission, Joint Research Center, Institute for Prospective Studies.
Leydesdorff, L. (2000). The triple helix: an evolutionary model of innovations. *Research Policy, 29*(2), 243-255.
Lim, K. (2004). The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981-1997). *Research Policy, 33*(2), 287-321.

Luker, W., & Lyons, D. (1997). Employment shifts in high technology industries, 1988-96. *Monthly Labor Review, June 1997*.

Lundvall, B. A., Johnson, B., Andersen, E. S., & Dalum, B. (2002). National systems of production, innovation and competence building. *Research Policy, 31*(2), 213-231.

Malkin, D. (2002). *Foresight and science and technology policy: An OECD perspective*. Paper presented at the The Role of Foresight in the Selection of Research Priorities, Seville, Spain.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy, 29*(3), 409-434.

Narin, F., Hamilton, K. S., & Olivastro, D. (1995). Linkage between agency supported research and patented industrial technology. *Research Evaluation, 5*(3), 183-187.

Nelson, R. R., & Rosenberg, N. (1993). Technical innovation and national systems. In R. R. Nelson (Ed.), *National Innovation Systems: A Comparative Analysis* (pp. 1-21.). New York: Oxford University Press.

Nelson, R. R., & Winter, S. G. (1977). In search of a useful theory of innovation. *Research Policy, 6*.

OECD. (1997a). *National Innovation Systems*. Paris, France: OECD. Online.

OECD. (1997b). *Oslo Manual. The Measurement of Scientific and Technological Activities, Proposed Guidelines for*
Collecting and Interpreting Technological Innovation Data.: http://www.oecd.org/pdf/M00018000/M00018312.pdf.

Pavitt, K. (1991). What makes basic research economically useful? *Research Policy, 20*, 109-119.

Porter, A. L., & Cunningham, S. W. (2005). *Tech Mining: Exploiting New Technologies for Competitive Advantage*. Hoboken N.J.: Wiley-Interscience.

Porter, A. L., Roessner, D., Jin, X.-Y., & Newman, N. C. (2002). Measuring national emerging technology capabilities. *Science and Public Policy, 29*(3), 189-200.

Riche, R., Hecker, D., & Burgan, J. (1983). High technology today and tomorrow: A small slice of the employment pie. *Monthly Labor Review, November 1983*, 50-58.

Rogers, E. M. (2003). *Diffusion of Innovations*. New York: Free Press.

Roy, S., Gevry, D., & Potenger, W. (2002). *Methodologies for trend detection in textual data mining*. Paper presented at the Second SIAM International Conference on Data Mining, Washington DC.

Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*.

Smalheiser, N. R. (2001). Predicting emerging technologies with the aid of text based data-mining. *Technovation, 21*, 689-693.

Small, H. (1999). Vizualizing science by citation mapping. *Journal of the American Society for Information science, 50*(9), 799-813.

Small, H. (2003). Paradigms, citations and maps of science: A personal history. *Journal of the American Society for Information science, 54*(5), 394-399.

Stokes, D. (1997). *Pasteur's Quadrant. Basic Science and Technological Innovation*. Washington, D.C.: The Brookings Institutions.

Teece, D. J. (2002). *Managing Intellectual Capital: Organizational, Strategic, and Policy Dimensions*: Oxford University Press.

Tijssen, R. J. W. (2002). Science dependence of technologies: evidence from inventions and their inventors. *Research Policy, 31*(4), 509-526.

Wagner, C. S., & Popper, S. W. (2003). Identifying critical, technologies in the United States: a review of the federal effort. *Journal of Forecasting, 22*(2-3), 113-128.

Walsh, S. T., & Linton, J. D. (2000). Infrastructure for Emergent Industries based on discontinuous innovations. *Engineering Management Journal, 12*(2), 23.

Watts, R. J., & Porter, A. L. (1997). Innovation forecasting. *Technological Forecasting and Social Change, 56*(1), 25-47.

APPENDIX

Summary of Data Sources Typically Used in *VantagePoint*

S & T Publications	**Patents**
• Engineering Village including EI Compendex and INSPEC • Web of Knowledge including science Citation Index and Social Science Citation • Chem Abstracts (CAS) • MEDLINE • NTIS • Pascal • RaDiUS • ResearchIndex	• MicroPatent • Delphion • PatentCafe.com • Questel-Orbit • WIPS • Derwent World Patent Index and Patent Citation • IFI CLAIMS

Source: (Porter and Cunningham, 2005)

Tech Mining Questions and Sample Indicators

Tech Mining Questions	**Measures and Innovation Indicators**
1. What emerging technologies merit our on-going attention?	• Scorecard measures • Trend plots in publication activity
2. What facets of this technology development are especially hot?	• PCA mapping of keyword clusters • 3-D trend charts for topics (component technologies)
3. What are new frontiers for this technology?	• Use list comparisons to ascertain new topics. • Use NLP on titles and abstracts to generate candidate new topics.
4. What are the component technologies that contribute importantly	• Map topical clusters (PCA). • Map class codes (publications).
5. How does this technological development fit with the technological landscape?	• White space maps. • 3-D surface maps.
6. What is driving this technology development?	• Score relative science base (% of patents citing R & D papers). • Relative publishing by academic, industry, and other organizations
7. What are key competing technologies?	• Scan of mentions in conjunction with the target technologies. • Compare maturation, drivers etc. of identified

	technologies with target.
8. How bright are the development prospects for this technology?	• Scorecard measures for rapid overview of multiple technologies. • Research activity over time.
9. What are the likely development pathways for this technology?	• Indicate velocity (publication rate and rate of change). • Mark status and project tech development prospects on an S-curve
10. What are the component technologies that contribute importantly?	• Map topical clusters (PCA). • Map class codes (publications)
11. Assess the maturation of the component technologies.	• List related technologies for expert screening. • Generate "hot spots" map
12. Identify technology fusion potential.	• Track over time – publication topic linkage patterns. • Map high-level topic clustering
13. Should we apply for particular patents relating to this technology?	• Present one-pager to facilitate expert risk assessment. • Display hot spots activity intersecting this patent.
14. Develop a technology product roadmap.	• Consolidate information on components and their maturation, technology development paths, production process development
15. Assess the maturation of systems in which to apply this technology.	• Application systems profile.
16. Which aspects (main topics) of this technology match application interests?	• Breakout main topic by publication and/or patent claim. • Breakout main topic by class codes.
17. What are opportunities in this emerging technology?	• Profile patent assignees (how many and how strong). • Indicate patent density.
18. What societal and market needs do this technology and its applications address?	• Scan mentions in publications. • Trends in needs mentioned.
19. What applications offer promise for this technology?	• Use NLP on claims to search for applications. • Develop an applications thesaurus to screen claims for applications.
20. What are the global opportunities?	• Geo-plot patent assignee concentrations • Geo-plot research publication by author nationality.
21. What is changing in the competitive environment?	• Indicate new entrants (first patents) • Indicate changing rate of entrance of new patentees.
22. Does this technology offer strong commercialization prospects?	• Gauge the technological infrastructure • Benchmark patent activity against comparable activity.
23. Assess the competitive environment.	• Chart recent slope for no. of class codes appearing annually. • Map dispersion of the technology.
24. Who are the available experts?	• Profile most prolific and most cited inventors not associated with a large company. • Profile most prolific and cited authors.

25. Which universities or research labs lead in this technology – overall or in particular aspects?	• Profile "Top N" publishing organizations in recent years • Map publishing of leading research organizations on 3-D surfaces, showing evolution over time.
26. What are the strengths and gaps within our own organization?	• Tabulate publication and/ or patent activity in related technologies • Map who collaborates with whom inside organization.
27. Which companies lead in particular aspects of this technology?	• Profile "Top-N" patenting companies by main topics (use class codes). • Concentration indicator (% of patents held by top companies).
28. How strong are the leading companies R&D teams?	• Map inventors (co-invention teaming; topical emphases).
29. Which companies lead in this technology?	• Graph patent citation distribution for most cited companies.
30. How do leading companies' development emphases compare to ours?	• Compare IPC or manual codes against organizations codes.
31. What other technological strengths does each leading company have?	• Map patent emphases
32. Characterize a company's IP relating to this technology.	• .Profile leading assignees.
33. What smaller companies or individuals have attractive IP relating to this technology?	• Profile assignees (how many and how strong) in a composite visualization
34. Who partners with whom?	• Identify co-assignees and co-authors, profile timing of collaborations.
35. Competitor profiling.	• Profile capabilities.
36. What companies should be placed on watch?	• Profile candidate companies
37. Who might be prospects to license the IP?	• Identify organizations with complementary technologies under development.
38. How entrepreneurial is the competitive environment?	• Indicate velocity (slope of new small business entrants in patenting).
39. Assess each key competitor.	• Chart competitors' recent vs. earlier publishing and patenting rates.

Source: Adapted from Porter and Cunningham (2005).